

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:11:25 ; Search time 3528.8 Seconds  
(without alignments)  
9927.764 Million cell updates/sec

Title: US-08-731-499-2  
Perfect score: 723  
Sequence: 1 TGAAGCTGTCATGTTACC.....TGAATGGAATGATGATTGCG 723

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_hvg.\*
- 3: gb\_in.\*
- 4: gb\_on.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	722.2	99.9	723	6 AR070325	Sequence 2 from patent US 5892010.
2	722.2	99.9	723	6 BD085726	Genes from the 20Q13 amplicon and their uses
3	714.8	98.9	163196	9 AC005220	Patent: US 5892010-A 2 06-APR-1999;
4	573.4	79.3	181313	2 AC021963	Location/Qualifiers
5	169	23.4	2808	6 C0726468	1. 723
6	167.6	23.2	349	6 BD060529	/organism="unknown"
7	167.6	23.2	2813	6 C0776693	/mol_type="unassigned DNA"
8	167.6	23.2	2813	6 AR371889	Query Match 99.9%; Score 722.2; DB 6; Length 723;
9	167.6	23.2	2813	9 AF041260	Best Local Similarity 100.0%; Pred. No. 6.8e-192; Mismatches 0; Indels 0; Gaps 0;
10	167.6	23.2	3066	6 BD085735	Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
11	137.8	19.1	168876	2 AC126535	Qy 1 TGAAGCTGTCATGTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATTATTCTCTGG 60
12	137.8	19.1	233233	2 AC094145	Db 1 TGAAGCTGTCATGTTACCGTCTCTAACGTTGGACTCTTAAGAAAATGATTATTCTCTGG 60
13	127.4	17.6	199019	10 AL928812	Qy 61 TTCTAGACAGGCCAAATGTAATTCACCTACGTGGCAGATTAAGAGAGTGGGCTTACTAG 120
14	127.4	17.6	235411	2 AC084066	Db 61 TTCTAGACAGGCCAAATGTAATTCACCTACGTGGCAGATTAAGAGAGTGGGCTTACTAG 120
15	86.6	12.0	2916	10 AY219233	Qy 121 ATTTGATTGGTATTGAGCATGCTCTGAATGACAGTCCCAAAAGGACCTCTTATCCGT 180
16	41.8	5.8	7218	6 I66494	Db 121 ATTTGATTGGTATTGAGCATGCTCTGAATGACAGTCCCAAAAGGACCTCTTATCCGT 180
17	41.8	5.8	125020	9 AP429315	Qy 181 TCCTCCCTTTGGGAAGGGCTTTTGGCCACTTCCATGTGCGCAGTTGAGCTTGGAA 240
18	41.4	5.7	5541	5 AB097138	Db 181 TCCTCCCTTTGGGAAGGGCTTTTGGCCACTTCCATGTGCGCAGTTGAGCTTGGAA 240
19	41.2	5.7	957	6 AX506150	Qy 241 ATTTGGTCGTTGTACAAACAATGATTACTTCTCCAAGATGTGCTGTGTGTAAGAAATGGTC 300

## ALIGNMENTS

RESULT 1	AR070325	723 bp	DNA	linear	PAT 18-FEB-2000
LOCUS	Sequence 2 from patent US 5892010.				
DEFINITION	AR070325				
ACCESSION	AR070325				
VERSION	AR070325.1	GI:7221213			
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 723)				
AUTHORS	Gray,J., Collins,C., Hwang,S.-i., Godfrey,T., Kowbel,D. and Rommens,J.				
TITLE	Genes from the 20Q13 amplicon and their uses				
JOURNAL	Patent: US 5892010-A 2 06-APR-1999;				
FEATURES	Location/Qualifiers				
source	1. 723				
ORIGIN	/organism="unknown"				
	/mol_type="unassigned DNA"				

20	41.2	5.7	1284	8	ATH243888	AJ243888 Arabidops
21	41.2	5.7	1296	8	AB032865	AB032865 Arabidops
22	41.2	5.7	1304	8	AT01010500	AJ010500 Arabidops
23	41.2	5.7	1356	8	AY084558	AY084558 Arabidops
24	41.2	5.7	110000	2	AC112592_0	AC112592 Rattus no
25	41.2	5.7	110000	2	AC112592_1	Continuation (2 of
26	40.6	5.6	104429	9	AP003787	AP003787 Homo sapi
27	40.6	5.6	165330	10	AC124987	AC124987 Mus muscu
28	40.6	5.6	172880	9	AC027419	AC027419 Homo sapi
29	40.2	5.6	136867	9	AC108019	AC108019 Homo sapi
30	40	5.5	88931	8	AP006686	AP006686 Lotus cor
31	40	5.5	193870	10	AL731674	AL731674 Mouse DNA
32	40	5.5	206467	2	AC133430	AC133430 Rattus no
33	40	5.5	219712	2	AC127196	AC127196 Rattus no
34	39.8	5.5	121688	9	AC092533	AC092533 Homo sapi
35	39.8	5.5	150396	2	AC025489	AC025489 Homo sapi
36	39.8	5.5	158709	2	AC083827	AC083827 Homo sapi
37	39.8	5.5	164739	2	BX682546	BX682546 Homo sapi
38	39.6	5.5	110000	2	AC146444_3	Continuation (4 of
39	39.6	5.5	167207	2	BX571817	BX571817 Danio rer
40	39.6	5.5	243657	2	AC124924	AC124924 Rattus no
41	39.6	5.5	264463	2	AC103132	AC103132 Rattus no
42	39.4	5.4	159254	2	AC084394	AC084394 Homo sapi
43	39.4	5.4	175320	2	AC139292	AC139292 Mus muscu
44	39.4	5.4	209895	9	AC074391	AC074391 Homo sapi
45	39.4	5.4	211130	2	AC098891	AC098891 Rattus no

Db 241 ATTGGTGGCTTTACAACTAAGCATTAATCTCTCAAGATGTGCCTGTGTGTAAGAAATGGTC 300  
QY 301 ATAGATTCAAACTGTAGCTACTATGTGGACAGGGGGCAGCAAGGACCCCACTTTGTAA 360  
Db 301 ATAGATTCAAACTGTAGCTACTATGTGGACAGGGGGCAGCAAGGACCCCACTTTGTAA 360  
QY 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTAACCTGGCAAAATAATCCAG 420  
Db 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTAACCTGGCAAAATAATCCAG 420  
QY 421 GTGGTGTGTAGTACCAGTAGAGATTATAAGTCCAGGAAGTAGAATCAGCCTTACAA 480  
Db 421 GTGGTGTGTAGTACCAGTAGAGATTATAAGTCCAGGAAGTAGAATCAGCCTTACAA 480  
QY 481 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAAAGA 540  
Db 481 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAAAGA 540  
QY 541 GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTTCTCAGACAAATGGTAAGCCCTTTA 600  
Db 541 GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTTCTCAGACAAATGGTAAGCCCTTTA 600  
QY 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660  
Db 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660  
QY 661 ACAGGTAGTAGGACATAAATGTAGCTGGTGTGTATGGAATGTAATGGAATATGGATT 720  
Db 661 ACAGGTAGTAGGACATAAATGTAGCTGGTGTGTATGGAATGTAATGGAATATGGATT 720  
QY 721 GCG 723  
Db 721 GCG 723

RESULT 2  
BD085726  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
PN JP 2001524802-A/2  
PD 04-DEC-2001  
PF 15-JUL-1997 JP 1998506264  
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR  
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI  
DAVID KOWBEL,  
PI JOHANNA ROMMENS  
PC C12N15/11,C12Q1/68,A61K48/00  
CC Description of Artificial Sequence:lb11  
PH Key Location/Qualifiers  
FT source 1..723  
FT /organism='Artificial Sequence'.  
FEATURES  
source  
1..723  
/organism="synthetic construct"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:32630"

Query Match 99.9%; Score 722.2; DB 6; Length 723;  
Best Local Similarity 100.0%; Pred. No. 6.8e-192;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGGAGCTGTCTATGTTTACCGTCTTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60  
Db 1 TGGAGCTGTCTATGTTTACCGTCTTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60  
QY 61 TTTCTAGACAGCCCAATGTAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAG 120  
Db 61 TTTCTAGACAGCCCAATGTAATTCACCTACGTGGCAGATTAAAGAGGTGGGCTTACTAG 120  
QY 121 ATTTGATTTGGGTATTTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTTTATCCGT 180  
Db 121 ATTTGATTTGGGTATTTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTTTATCCGT 180  
QY 181 TCTTCCCTTTGGGAGGGCTTTTGGCACTTCCATGTCATATGTGTCAGTTGAGCTTGGAA 240  
Db 181 TCTTCCCTTTGGGAGGGCTTTTGGCACTTCCATGTCATATGTGTCAGTTGAGCTTGGAA 240  
QY 241 ATTTGATTTGGGTATTTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTTTATCCGT 300  
Db 241 ATTTGATTTGGGTATTTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTTTATCCGT 300  
QY 301 ATAGATTCAAACTGTAGCTACTATGTGGACAGGGGGCAGCAAGGACCCCACTTTGTAA 360  
Db 301 ATAGATTCAAACTGTAGCTACTATGTGGACAGGGGGCAGCAAGGACCCCACTTTGTAA 360  
QY 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTAACCTGGCAAAATAATCCAG 420  
Db 361 AACATGTTTGGGGGAATGTTTTGTTTTTCAATTTTCTTATTAACCTGGCAAAATAATCCAG 420  
QY 421 GTGGTGTGTAGTACCAGTAGAGATTATAAGTCCAGGAAGTAGAATCAGCCTTACAA 480  
Db 421 GTGGTGTGTAGTACCAGTAGAGATTATAAGTCCAGGAAGTAGAATCAGCCTTACAA 480  
QY 481 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAAAGA 540  
Db 481 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACWGAAGCGAAACTCAAAAGA 540  
QY 541 GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTTCTCAGACAAATGGTAAGCCCTTTA 600  
Db 541 GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTTCTCAGACAAATGGTAAGCCCTTTA 600  
QY 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660  
Db 601 CTTCCAGTATAGAAACCTAAGATACCTAGACGGGCTTTTGGGAACAATGGGCTCATGCC 660  
QY 661 ACAGGTAGTAGGACATAAATGTAGCTGGTGTGTATGGAATGTAATGGAATATGGATT 720  
Db 661 ACAGGTAGTAGGACATAAATGTAGCTGGTGTGTATGGAATGTAATGGAATATGGATT 720  
QY 721 GCG 723  
Db 721 GCG 723  
RESULT 3  
AC005220  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
AC005220  
Homo sapiens chromosome 20, BAC clone 99 (LBNL H80), complete  
sequence.  
AC005220  
AC005220.1 GI:3282159  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 163196)  
Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,  
Connolly,K.S., Gunning,K.M., Kadner,K., Miguel,F., Miller,C.,  
Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.

```

TITLE      Sequencing of human chromosome 20
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 163196)
AUTHORS    Ricke,D.O.
TITLE      Large Scale Sequence Analysis and Annotation with the Sequence
AUTHORS     Comparison Analysis (SCAN) System
JOURNAL    Unpublished
REFERENCE  3 (bases 1 to 163196)
AUTHORS     Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,
              Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Miguel,T.,
              Piluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.
TITLE      Direct Submission
JOURNAL    Submitted (01-JUL-1998) Human Genome Center, DOE Joint Genome
              Institute, Lawrence Berkeley National Laboratory, MS 74-157,
              Berkeley, CA 94720, U.S.A.
COMMENT     Sequence submitted by:
FEATURES   Location/Qualifiers
            source
              1..163196
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /chromosome="20"
                /map="20"
                /clone="BAC 99"
                /note="IBNL H80"
            repeat_region
              4588..4709
                /rpt_family="H80"
            repeat_region
              6685..6813
                /rpt_family="MIR"
            repeat_region
                /rpt_family="trigger1"
            repeat_region
                /complement(6838..7220)
            repeat_region
              7175..7314
                /rpt_family="L1"
            repeat_region
              7322..7441
                /rpt_family="trigger1"
            repeat_region
                /complement(8695..9257)
            repeat_region
              10121..10350
                /rpt_family="MLT2B2"
            repeat_region
              11057..11357
                /rpt_family="Alu"
            repeat_region
              13431..13702
                /rpt_family="Alu"
            repeat_region
              15911..15962
                /rpt_family="MIR"
            misc_feature
              16029..16484
                /note="99% identity B54864 BAC end sequence"
            repeat_region
                /complement(16891..17025)
            repeat_region
                /rpt_family="Alu"
            repeat_region
                /complement(17221..17682)
            repeat_region
                /rpt_family="L1"
            repeat_region
                /complement(18500..18756)
            misc_feature
                /rpt_family="L1"
            repeat_region
                /complement(18909..19161)
            repeat_region
              19640..19943
                /note="GRAIL 2 excellent exon, frame 1"
            repeat_region
                /rpt_family="Alu"
            repeat_region
                /complement(22025..22314)
            misc_feature
                /rpt_family="Alu"
            repeat_region
                /complement(22903..23020)
            repeat_region
                /note="GRAIL 2 excellent exon, frame 1"
            repeat_region
                /complement(23247..23567)
            misc_feature
                /rpt_family="Alu"
            repeat_region
                /complement(24411..24514)
            repeat_region
                /note="GRAIL 2 excellent exon, frame 0"
            repeat_region
              25506..25525
                /note="(T) 20"
            repeat_region
                /rpt_type=tandem
            repeat_region
                /rpt_unit="t"
            repeat_region
                /complement(25538..25679)
            repeat_region
                /rpt_family="MIR"
            repeat_region
                /complement(26920..27577)
            repeat_region
                /rpt_family="L1"
            repeat_region
                /complement(28412..28563)
            repeat_region
                /rpt_family="MER20"
            repeat_region
              28734..29019
                /rpt_family="Alu"
            repeat_region
              30266..30455
                /rpt_family="THE1"
            misc_feature
              30789..30888
                /note="GRAIL 2 excellent exon, frame 2"
            misc_feature
              31841..31972
                /note="GRAIL 2 excellent exon, frame 0"
            repeat_region
              32270..32290
                /note="(T) 21"
            repeat_region
                /rpt_type=tandem
            repeat_region
                /rpt_unit="t"
            misc_feature
              32702..32860
                /note="GRAIL 2 excellent exon, frame 1"
            misc_feature
              33268..33474
                /note="GRAIL 2 excellent exon, frame 0"
            repeat_region
              33836..33863
                /note="(GT) 14"
            repeat_region
                /rpt_type=tandem
            repeat_region
                /rpt_unit="gt"
            misc_feature
              35158..35265
                /note="GRAIL 2 excellent exon, frame 0"
            repeat_region
                /complement(41855..42137)
            repeat_region
                /rpt_family="Alu"
            misc_feature
              43253..43465
                /note="GRAIL 2 excellent exon, frame 1"
            repeat_region
                /complement(50487..50600)
            repeat_region
                /rpt_family="L1"
            repeat_region
              50605..50855
                /rpt_family="L1"
            repeat_region
              51461..51822
                /rpt_family="THE1"
            repeat_region
                /complement(54343..54856)
            repeat_region
                /rpt_family="L1"
            repeat_region
                /complement(55490..55796)
            repeat_region
              55900..56027
                /rpt_family="MIR"
            repeat_region
                /complement(61007..61096)
            repeat_region
                /note="GRAIL 2 excellent exon, frame 1"
            repeat_region
                /standard_name="af74a01"
            repeat_region
                /note="100% identity dbEST:AA644590"
            repeat_region
                /complement(62247..62746)
            repeat_region
                /rpt_family="L1"
            repeat_region
                /complement(65258..65541)
            repeat_region
              65587..65882
                /rpt_family="Alu"
            repeat_region
                /rpt_family="Alu"
            repeat_region
              68256..68282
                /note="(T) 27"
            repeat_region
                /rpt_type=tandem
            repeat_region
                /rpt_unit="t"
            repeat_region
              69077..69495
                /rpt_family="Alu"
            repeat_region
              69537..69618
                /rpt_family="Alu"
            repeat_region
                /complement(70542..70721)
            repeat_region
                /rpt_family="MERS"
            repeat_region
              72072..72255
                /rpt_family="MIR"
            repeat_region
                /complement(72260..72340)
            repeat_region
                /rpt_family="MLT1"
            repeat_region
              74417..74456
                /note="(CT) 20"
            repeat_region
                /rpt_type=tandem
            repeat_region
                /rpt_unit="ct"
            repeat_region
              75057..75320
                /note="GRAIL 2 excellent exon, frame 2"
            repeat_region
                /complement(75199..75323,84601..84746)
            repeat_region
                /complement(75199..75323,84601..84746)

```







```
QY 481 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCGAAACTCAAAAGA 540
|||||
Db 37065 ACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCGAAACTCAAAAGA 37006
|||||
QY 541 GAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTC 579
|||||
Db 37005 GAAGAAAGCAAAACCAAGAACCTCTCTGATGGCGTTTCTC 36967
|||||

RESULT 5
CQ726468 2808 bp DNA linear PAT 03-FEB-2004
LOCUS
DEFINITION Sequence 12402 from Patent WO02068579.
ACCESSION CQ726468
VERSION CQ726468.1 GI:42289602
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Venter,C.J., Adams,M.C., Li,P.W. and Myers,E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 12402 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source Location/Qualifiers
1..2808
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 23.4%; Score 169; DB 6; Length 2808;
Best Local Similarity 95.0%; Pred. No. 2.6e-36;
Matches 172; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 408 CAAATAATCCAGGTGTTGAGTCACCAAGAGAGATTATAAAGTCCCAAGGAAGTAGA 467
|||||
Db 1279 CCAAAAGGACCTGTGGTGTGAGTCACCAAGAGAGATTATAAAGTCCCAAGGAAGTAGA 1338
|||||
QY 468 ATCAGCGTTTACAAACAGTGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGC 527
|||||
Db 1339 ATCAGCGTTTACAAACAGTGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGC 1398
|||||
QY 528 GAAACTCAAAAGAGAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAAT 587
|||||
Db 1399 GAAACTCAAAAGAGAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAAT 1458
|||||
QY 588 G 588
Db 1459 G 1459

RESULT 6
BD060529 349 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION Secreted expressed sequence tags (seSTs).
ACCESSION BD060529
VERSION BD060529.1 GI:22606135
KEYWORDS Zea mays
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 349)
REFERENCE
AUTHORS Jacobs,K., Mccoy,J.M., Lavallie,E.R., Racie,L.A., Merberg,D.,
Treacy,M., Spaulding,V. and Agostino,M.J.
TITLE Secreted expressed sequence tags (seSTs)
JOURNAL Patent: JP 2001518793-A 889 16-OCT-2001;

GENETICS INSTITUTE INC
PN JP 2001518793-A/889
PD 16-OCT-2001
PF 10-APR-1998 JP 1998543070
PR 10-APR-1997 US 08/837312
PI KENNETH JACOBS, JOHN M MCCOY, EDWARD R LAVALLIE, LISA A RACIE, PI
DAVID MERBERG,
PI MAURICE TREACY, VIKKI SPAULDING, MICHAEL J AGOSTINO PC
C12N15/12, C12N5/10, C07K14/47, C12Q1/68, A61K38/17 CC Strandedness:
Double;
CC Topology: Linear; Location/Qualifiers.
FH Key Location/Qualifiers
1..349
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"

FEATURES
source Location/Qualifiers
1..349
/organism="Zea mays"
/mol_type="genomic DNA"
/db_xref="taxon:4577"

ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 349;
Best Local Similarity 98.2%; Pred. No. 7.1e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 478
|||||
Db 69 ATGTGTGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 128
|||||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538
|||||
Db 129 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 188
|||||
QY 539 GAGAAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
|||||
Db 189 GAGAAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 238
|||||

RESULT 7
CQ776693 2813 bp DNA linear PAT 11-MAR-2004
LOCUS
DEFINITION Sequence 379 from Patent EPI394274.
ACCESSION CQ776693
VERSION CQ776693.1 GI:45380083
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Ohtani,N., Sugita,Y., Yamaya,M., Kubo,H., Nagai,H. and Izuwara,K.
TITLE Methods of testing for bronchial asthma or chronic obstructive
pulmonary disease
JOURNAL Patent: EP 1394274-A 379 03-MAR-2004;
Genox Research, Inc. (JP)
FEATURES
source Location/Qualifiers
1..2813
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 2813;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 478
|||||
Db 1298 ATGTGTGTGTGAGTCACCAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1357
|||||
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538
|||||
Db 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1417
|||||
QY 539 GAGAAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
|||||
```

```
Db 1418 GAGAGAGAGGACCAAGAACCTCTCTGATGGCGTTCTCAGACAAATG 1467
|||||
RESULT 8
AR371889 2813 bp DNA linear PAT 12-SEP-2003
LOCUS
DEFINITION Sequence 3 from patent US 6395544.
ACCESSION AR371889
VERSION AR371889.1 GI:34608999
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2813)
AUTHORS Cowser, L.M. and Freier, S.M.
TITLE Antisense modulation of BCAS1 expression
JOURNAL Patent: US 6395544-A 3 28-MAY-2002;
FEATURES
source Location/Qualifiers
1.2813
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 2813;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 419 AGGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAGTGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAGTGAATCAGCCTTAC 1357
Qy 479 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGACACCCACGAGGAGGAACTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGACACCCACGAGGAGGAACTCAAAA 1417
Qy 539 GAGAGAGAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
Db 1418 GAGAGAGAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1467
|||||
RESULT 9
AF041260 2813 bp mRNA linear PRI 22-JUL-1998
LOCUS
DEFINITION Homo sapiens AIB1 (AIB1) mRNA, complete cds.
ACCESSION AF041260
VERSION AF041260.1 GI:3335398
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2813)
AUTHORS Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M.,
Hwang, S.-I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K.,
Jay, K.E., Froula, J., Cloutier, T., Kuo, W.-L., Yaswen, P., Dairkee, S.,
Giovannola, J., Hutchinson, G.B., Isola, J., Kallioniemi, O.-P.,
Palazzolo, M., Martin, C., Ericsson, C., Pinkel, D., Albertson, D.,
Li, W.-B. and Gray, J.W.
TITLE Positional cloning of ZNF217 and NAB1: genes amplified at 20q13.2
and overexpressed in breast carcinoma
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8703-8708 (1998)
MEDLINE 98337979
PUBMED 9671742
REFERENCE 2 (bases 1 to 2813)
AUTHORS Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M.,
Hwang, S.-I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K.,
Jay, K.E., Froula, J., Cloutier, T., Kuo, W.-L., Yaswen, P., Dairkee, S.,
Giovannola, J., Hutchinson, G.B., Isola, J., Kallioniemi, O.-P.,
Palazzolo, M., Martin, C., Ericsson, C., Pinkel, D., Albertson, D.,
Li, W.-B. and Gray, J.W.
TITLE Direct Submission
JOURNAL Submitted (07-JAN-1998) Life Sciences, Lawrence Berkeley National
```

```
FEATURES
source Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA
Location/Qualifiers
1.2813
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.2-q13.3"
gene 1.2813
/feature="AIB1"
CDS 118..1872
/feature="AIB1"
/notes="amplified and overexpressed in breast cancer"
/codon_start=1
/product="AIB1"
/protein_id="AAC39896.1"
/db_xref="GI:3335398"
/tranlation="MGQMSVQPVQVEDEPEAEYQDNASALNGVPVVVSTHTVQH
LEEVDLGISVKTNDVATSSPETTEISAVADANGKNGLEAKPEAPAKSRFFLMLSRP
VPGRTGDAADSLGKLVSNKAPKDPSESWTLPVAGPGQDTKTPGHPAQ
DKVLSAARDPTLLPETGGAGRAPKPSFDFKFLDKGQKVPQSGQAKRA
EHQDKVDEVPGLSQSDVDVAGKDIIVGKEKEGELGTADCSVPDPEGLTAKDDSO
AAAIANNNSIMSFFKTLVSPNKAETKKDPEDTGAESPTTSADLKSKDKANFTSOETQ
GAGKNSKGNPSGHTQSVTPPEAKEGTSPTSLPLKLFHKKSVKEDSVTGAIE
ENVVCSPEVILIKSEVESALQTVLDNEGDAPEPTAKLKEESKPTSLMAFLROM
SVKGDGITHSEINGKSSQTSDEKTTTPPEPTGAPQKQKGGSKOKKSAAE
MNFQKNQKQAKPEAQCETQATVDTNSLQNGDKLQKPEKQKQSLGSGFKGLGPKRML
DAQVQTDVPSIGPVKPK"
ORIGIN
Query Match 23.2%; Score 167.6; DB 9; Length 2813;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 419 AGGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAGTGAATCAGCCTTAC 478
Db 1298 ATGTGGTGTGTGAGTCACAGGAGATTTAAAGTCCAGGAGTGAATCAGCCTTAC 1357
Qy 479 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGACACCCACGAGGAGGAACTCAAAA 538
Db 1358 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGACACCCACGAGGAGGAACTCAAAA 1417
Qy 539 GAGAGAGAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
Db 1418 GAGAGAGAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1467
|||||
RESULT 10
BD085735 3066 bp DNA linear PAT 27-AUG-2002
LOCUS
DEFINITION Genes from the 20q13 amplicon and their uses.
ACCESSION BD085735
VERSION BD085735.1 GI:22631345
KEYWORDS JP 2001524802-A/11.
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1 (bases 1 to 3066)
AUTHORS Gray, J.W., Collins, C.C., Hwang, S.I., Godfrey, T., Kowbel, D. and
Rommens, J.
TITLE Genes from the 20q13 amplicon and their uses
JOURNAL Patent: JP 2001524802-A 11 04-DEC-2001;
COMMENT THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
OS Artificial Sequence
PN JP 2001524802-A/11
PD 04-DEC-2001
PR 15-JUL-1997 JP 1998506264
PR 15-JUL-1996 US 08/680395, 16-OCT-1996 US 08/731499 PR
17-JAN-1997 US 08/785532
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI
DAVID KOWBEL,
PI JOHANNA ROMMENS
PC C12N15/11.C12Q1/68.A61K48/00
```

```

CC Description of Artificial Sequence:lb1
FH Key 1. .3066
FT source Location/Qualifiers
FT /organism='Artificial Sequence'
FT /organism='synthetic construct'
FT /mol_type='genomic DNA'
FT /db_xref='taxon:32630'

FEATURES
source
ORIGIN
Query Match 23.2%; Score 167.6; DB 6; Length 3066;
Best Local Similarity 98.2%; Pred. No. 6.4e-36;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 419 AGTGTGTGTGAGTCACCAAGTAGAGATTTATAAGTCCAAAGGAGTAGAATCAGCCTTAC 478
DB 1433 ATGTGTGTGTGAGTCACCAAGTAGAGATTTATAAGTCCAAAGGAGTAGAATCAGCCTTAC 1492

OY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCCAAGCAAGCAAACTCAAAA 538
DB 1493 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCCAAGCAAGCAAACTCAAAA 1552

OY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
DB 1553 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 1602

RESULT 11
AC126535/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-408J13, *** SEQUENCING IN PROGRESS
*** 3 unordered pieces.
AC126535
VERSION V.1
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 168876)
Muzny,D.Marie, Metzker,M.Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Gantä,R., Garcia,A., Garner,T., Garza,M.,
Guebrgeorgis,E., Geer,K., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulvyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,B., Kovar,C.,
Kowitz,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewari,L., Louisedge,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,

```

```

Nwackeleme,O., Okwuonu,G., Olarnpunaagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfanckoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzio,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Saverty,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 168876)
Worley,K.C.
Direct Submission
Submitted (06-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168876)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (19-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 19, 2002 this sequence version replaced gi:21703469.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the
sequence may extend beyond the ends of the clone and there may be
contigs that consist entirely of whole genome shotgun sequence
reads. Both end sequences and whole genome shotgun sequence only
contigs will be indicated in the feature table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYPD
Center clone name: CH230-408J13
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 151329 bases at least Q40
Consensus quality: 153110 bases at least Q30
Consensus quality: 154290 bases at least Q20
Estimated insert size: 162792; sum-of-contigs estimation
Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 12922: contig of 12922 bp in length
* 12923 13022: gap of unknown length
* 13023 155459: contig of 142437 bp in length
* 155460 155559: gap of unknown length
* 155560 168876: contig of 13317 bp in length.
* Location/Qualifiers
* 1 .168876
FEATURES
source

```

```

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-408013"
11757..12922
/notes="wgs_contig"
13023..14162
/notes="wgs_contig"
15743..16838
/notes="wgs_contig"

```

## ORIGIN

```

Query Match      19.1%; Score 137.8; DB 2; Length 168876;
Best Local Similarity 73.8%; Pred. NO. 1.3e-27;
Matches 172; Conservative 2; Mismatches 59; Indels 0; Gaps 0;

Qy 369 TTGGGGGAATGTTTGTGTTTTCATTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTG 428
Db 143490 TGGGGTGTGTTTGTGTTTCTTATTACCTGGCAAAACACTCTAGGCGGTGTG 143431

Qy 429 TGAGTCCACGTAGAGATTAAAGTCCAGGAAGTAGAATCAGCTTACAAAACAGTGGGA 488
Db 143430 TGAGCCACCGTAGAGACTGTAAGGCTTGAGGAAGTAGAATCCACCTTACAAACCGTGGGA 143371

Qy 489 CCTCAACGAGAGATGCTGCACCTGACCCACGACGCAAGCACTCAAAAGAGAAAG 548
Db 143370 TCTCACTGAGAAAGAAACCCAGACTGAGCCACAGATGTGAAGTCAAAAGAGAAAGCAA 143311

Qy 549 CAACACCAAGAACCTCTCGATGTCGTTCTCAGACAAATGTAAGCCCTTAC 601
Db 143310 ACCCGGAGGACCCCTCGATGTCGTTCTCAGACAAATGTAAGCCACTTCC 143258

```

## RESULT 12

```

AC094145/c. 233233 bp DNA linear HTG 09-MAY-2003
LOCUS Rattus norvegicus clone CH230-2L13, WORKING DRAFT SEQUENCE, 6
DEFINITION unordered pieces.

```

## ACCESSION

```

AC094145
AC094145.6 GI:30467887

```

## KEYWORDS

```

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.

```

## SOURCE

```

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

```

## REFERENCE

```

1 (bases 1 to 233233)
Muzny D.Marie., Metzker M.Lee., Abramson S., Adams C., Alder J.,
Allen C., Allen H., Alsbrooks S., Amin A., Anguiano D.,
Anyalebechi V., Aoyagi A., Ayodeji M., Baca E., Baden H.,
Baldwin D., Bandaranaike D., Barber M., Barnstead M., Benahmed F.,
Biswal K., Blair J., Blankenburg K., Blyth P., Brown M.,
Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,
Chacko J., Chavez D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
Cleveland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
Delgado O., Denison S., Deramo C., Ding Y., Dinh H., Divya K.,
Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
Fraser C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
Gebregregis E., Geer K., Gill R., Grady M., Guerra R., Guevara W.,
Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,
Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,
Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,
Hollins B., Howells S., Hulyk S., Hume J., Idlebird D., Jackson A.,
Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolivet A.,
Karpathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
Kovis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
Lorensuewa L., Loulisedge H., Lozano R.J., Lu X., Ma J.,
Maheshwari M., Mahindartne M., Mahmoud M., Malloy K., Mangum A.,

```

Mangum B., Mapua P., Martin K., Martin R., Martinez E.,  
Mawhney S., McLeod M.P., McNeill T.Z., Meenen E.,  
Milosavljevic A., Miner G., Minja E., Montemayor J., Moore S.,  
Morgan M., Morris K., Morris S., Munidasa M., Murphy M., Nair L.,  
Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,  
Nwaokemele O., Okwuonu G., Olarnpungoon A., Pal S., Parks K.,  
Pasternak S., Paul H., Perez A., Perez L., Pfannkuch C.,  
Popper F., Poindexter A., Popovic D., Primus E., Pu L.-L.,  
Puzo M., Quiroz J., Rachlin E., Reeves K., Regier M.A., Reigh R.,  
Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,  
Rives C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,  
Sanders W., Savery G., Scherer S., Scott G., Shateman S., Shen H.,  
Shetty J., Shvartsbeyn A., Sisson I., Sitter C.D., Smajda D.,  
Sneed A., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,  
Steimle M., Strong R., Sutton A., Svatek A., Taber P., Taylor C.,  
Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,  
Valas R., Vera V., Villalana D., Waldron L., Walker B., Wang J.,  
Wang O., Wang S., Warren J., Warren R., Wei X., White F.,  
Williams G., Willson R., Wleczyk R., Wooden H., Worley K.,  
Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,  
Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von  
Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,  
Weinstock G. and Gibbs R.A.

## Direct Submission

Unpublished

2 (bases 1 to 233233)

Worley K.C.

## Direct Submission

Submitted (17-SEP-2001)

Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 233233)

Rat Genome Sequencing Consortium.

## Direct Submission

Submitted (09-MAY-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On May 9, 2003 this sequence version replaced gi:24942806.

The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

## Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GABR

Center clone name: CH230-2L13

----- Summary Statistics

Assembly program: Atlas;

Consensus quality: 207295 bases at least Q40

Consensus quality: 210461 bases at least Q30

Consensus quality: 212871 bases at least Q20

Estimated insert size: 217628; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 6 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 89147: contig of 89147 bp in length  
 \* 89148 89247: gap of unknown length  
 \* 89248 99513: contig of 10266 bp in length  
 \* 99514 99613: gap of unknown length  
 \* 99614 113910: contig of 14297 bp in length  
 \* 113911 114010: gap of unknown length  
 \* 114011 230487: contig of 116477 bp in length  
 \* 230488 230587: gap of unknown length  
 \* 230588 232001: contig of 1414 bp in length  
 \* 232002 232102: gap of unknown length  
 \* 232102 232323: contig of 1132 bp in length.

## FEATURES

source

1. .232323  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-2L13"

misc\_feature

1. .1412  
 /note="wgs\_contig"

## ORIGIN

Query Match 19.1%; Score 137.8; DB 2; Length 232323;  
 Best Local Similarity 73.8%; Pred. No. 1.2e-27;  
 Matches 172; Conservative 2; Mismatches 59; Indels 0; Gaps 0;  
 QY 369 TTGGGGGAATGTTTCTTATTTCTTATTTACCTGGCAAAATATCAGGTGTGTG 428  
 Db 77599 TGGGGTGTGTTTGTGTTTGTGTTTCTTATTTACCTGGCAAACTCTAGCGGTGTG 77540  
 QY 429 TGAGTCACAGTAGATTAAGTCCAAAGGAGTAGATCAGCTTACAACTGCA 488  
 Db 77539 TGAGCCACAGTAGAGACTGTAAGGCTTGAGGAAGTAGATCACCCTTACAACTGGA 77480  
 QY 489 CCTCAACGAGGAGAGTCTGCACCTGAACCCACGAGGCAAACTCAAAAGAGAGAAAG 548  
 Db 77479 TCTCACTGAGAAAGAACCCAGACTGAGCCACAGATGTGAAGTCAAAAGAGAAAGCA 77420  
 QY 549 CAACCAAGAACTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTAC 601  
 Db 77419 ACCCGGAGGAGCCCTCTGATGCGTTTCTCAGACAAATGGTAAGCCACTTC 77367

## RESULT 13

AL928812/c

LOCUS Mouse DNA sequence from clone RP23-321B23 on chromosome 2, complete  
 DEFINITION sequence.

AL928812.11 GI:31335541

VERSION HTG.

KEYWORDS

SOURCE Mus musculus (house mouse)

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 199019)

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (30-MAY-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Jun 2, 2003 this sequence version replaced gi:31096514.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.

----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk

-----  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/projects/C\_elegans/wormpep RP23-321B23 is  
 from the RPCI-23 Mouse BAC Library  
 constructed by the group of Pieter de Jong.

For further details see http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6.

## FEATURES

source

Location/Qualifiers  
 1. .199019  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-321B23"  
 /clone\_lib="RPCI-23"

## ORIGIN

Query Match 17.6%; Score 127.4; DB 10; Length 199019;  
 Best Local Similarity 74.7%; Pred. No. 1.1e-24;  
 Matches 171; Conservative 2; Mismatches 53; Indels 3; Gaps 1;  
 QY 371 GGGGGAATCTTTGTTTCTTATTTACCTGGCAAAATATCAGGTGTGTG 430  
 Db 165827 GGGTGTGTTTGTGTTTGTGTTTCTTATTTACCTGGCAAAATATCAGGTGTGTG 165768  
 QY 431 AGTCACAGTAGAGATTATAAGTCCAAAGAGTAGAATCAGCTTACAACTGAGCC 490  
 Db 165767 AGTCACAGTAGAGACCGTAAGCTTGAGGAAGTAGAATCCAGCTTACAACTGAGATC 165708  
 QY 491 TCACGAGAGAGATGCTGCACCTGACCCACGAGCAAACTCAAAAGAGAGAAAGCA 550  
 Db 165707 TCAGTGA---AGAGACCCAGCCTGAACCCACAGACGTAAGTCAAAAGAGAAAGCAAC 165651  
 QY 551 AACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTT 599  
 Db 165650 CCGGGAAGACCCCTCTGATGGGTTTCTCAGACAAATGGTAAGCCACTT 165602

## RESULT 14

AC084066/c

LOCUS

DEFINITION

Mus musculus clone RP23-321D1, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 29

unordered pieces.

AC084066

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 235411)

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (12-OCT-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: <http://www.jgi.doe.gov>  
-----  
Project Information  
Center Project Name: 2351294  
Center clone name: RPCI-23\_321D1  
-----

Summary Statistics

Consensus quality: 214207 bases at least Q40  
Consensus quality: 223053 bases at least Q30  
Consensus quality: 225208 bases at least Q20  
Estimated insert size: 200000; pulse field gel estimation  
Estimated insert size: 232611; sum-of-contigs estimation  
Quality coverage: 11.56 in Q20 bases; pulse field gel estimation  
Quality coverage: 9.94 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 29 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1210: contig of 1210 bp in length  
\* 1211 1310: gap of unknown length  
\* 1311 2240: contig of 1930 bp in length  
\* 3241 3340: gap of unknown length  
\* 3341 4504: contig of 1164 bp in length  
\* 4505 4604: gap of unknown length  
\* 4605 6172: contig of 1568 bp in length  
\* 6173 6272: gap of unknown length  
\* 6273 7354: contig of 1082 bp in length  
\* 7355 7454: gap of unknown length  
\* 7455 8625: contig of 1171 bp in length  
\* 8626 8725: gap of unknown length  
\* 8726 10114: contig of 1389 bp in length  
\* 10115 10214: gap of unknown length  
\* 10215 12091: contig of 1877 bp in length  
\* 12092 12191: gap of unknown length  
\* 12192 14113: contig of 1922 bp in length  
\* 14114 14213: gap of unknown length  
\* 14214 15286: contig of 1073 bp in length  
\* 15287 15386: gap of unknown length  
\* 15387 17839: contig of 2453 bp in length  
\* 17840 17939: gap of unknown length  
\* 17940 20266: contig of 2327 bp in length  
\* 20267 20366: gap of unknown length  
\* 20367 22001: contig of 1635 bp in length  
\* 22002 22101: gap of unknown length  
\* 22102 24307: contig of 2206 bp in length  
\* 24308 24408: gap of unknown length  
\* 24408 26405: contig of 1998 bp in length  
\* 26406 26505: gap of unknown length  
\* 26506 29691: contig of 3186 bp in length  
\* 29692 29791: gap of unknown length  
\* 29792 34705: contig of 4914 bp in length  
\* 34706 34805: gap of unknown length  
\* 34806 39749: contig of 4944 bp in length  
\* 39750 39849: gap of unknown length  
\* 39850 45296: contig of 5447 bp in length  
\* 45297 45397: gap of unknown length  
\* 45397 51476: contig of 6080 bp in length  
\* 51477 51576: gap of unknown length  
\* 51577 59008: contig of 7432 bp in length  
\* 59009 59108: gap of unknown length  
\* 59109 66218: contig of 7110 bp in length  
\* 66219 66318: gap of unknown length  
\* 66319 76778: contig of 10460 bp in length

FEATURES

source  
1. .235411  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/clone="RP23-321D1"  
/clone\_lib="RPCI mouse BAC library 23"

ORIGIN

Query Match 17.6%; Score 127.4; DB 2; Length 235411;  
Best Local Similarity 74.7%; Pred. No. 1e-24;  
Matches 171; Conservative 2; Mismatches 53; Indels 3; Gaps 1;  
Qy 371 GGGGAATGTTTCTTCTTCAATTTCTTATTACCTGCGCAAAATATCCAGGTGTGTG 430  
Db 225082 GGGTGTGTTTGTGTTTGTGTTTCTTATTACCTGCGCAAAATATCTAGGCGGTGTG 225023  
Qy 431 AGTCACAGTAGAGATTATAAGTCCAAAGGAGTAGAATCAGCCTTACAAAACAGTGGACC 490  
Db 225022 AGTCACAGTAGAGACCGTAGGCTTGAGGAGTAGAATCAGCCTTACAAAACAGTGGATC 224963  
Qy 491 TCAACGAAGGAGATGCTGCACCTGAACCCACGAAAGCGAAACTCAAAAGAGAAGAAAGCA 550  
Db 224962 TCAGTGA---AGAGACCCAGCCTCAACCCACAGACGTAAAAGTCAAGAGAAGAAAC 224906  
Qy 551 AACCAAGAACTCTCTGATGCGTTTCTCAGACAAATGGTAAGCCCTT 599  
Db 224905 CCCGAAGACCCCTCTGATGCGTTTCTCAGACAAATGGTAAGCCACTT 224857

RESULT 15

AY219233 2916 bp mRNA linear ROD 20-OCT-2003  
LOCUS Mus musculus novel amplified in breast cancer-1 mRNA, complete cds.  
DEFINITION  
ACCESSION AY219233  
VERSION AY219233.1 GI:29648617  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 2916)  
AUTHORS Beardsley, D.I., Kowbel, D., Lataxes, T.A., Mannino, J.M., Xin, H.,  
Kim, W.J., Collins, C. and Brown, K.D.  
TITLE Characterization of the novel amplified in breast cancer-1 (NABCL1)  
gene product  
JOURNAL Exp. Cell Res. 290 (2), 402-413 (2003)  
MEDLINE 22930202  
PUBMED 14567997  
REFERENCE 2 (bases 1 to 2916)  
AUTHORS Brown, K.  
TITLE Direct Submission  
JOURNAL Submitted (14-JAN-2003) Biochem. and Molec. Biology, LSU Health  
Sciences Center, 1901 Perdido Street, New Orleans, LA 70112, USA  
FEATURES  
source  
1. .2916  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
89..1990  
/note="NABCL1; AIBCL1; BCAS1"

CDS

```
/codon start=1
/product="novel amplified in breast cancer-1"
/protein_id="A08801.1"
/db_xref="GI:29648618"
/translation="MGNQMSVPLRPGDQHDGADTCVKVTSNECVQNGNPVVLSTRV
IQHYEVDLIGISSKDNVATSPFKMEQAQVGDAGSKNLGKEAKTKAPAAHFFLTL
SRPVFRPGDQGTDSAAERFDVSPAAADPNKSEHGALFVAAAPGAQADKTPCCP
EAKQOTLPATGPSPESQAEAPQADKDFGLNRFKLDKGRESAPVNSQPKKAGG
SDPQATAPAVPGNPHGVSAgedIvdSEQRQDVTLSYSPGDPEVPGTTKEDPQ
VVDTTSSSIIMSFKTLVSPNKTETKDPEDTKATKADSVCDGHAAGKMSQTQKS
KKRLDSPRLGUSFRKLRKDTENSPTTSANLKDKNFTQETRGKTKATKSCSP
PPSPFTSGRDSGKEKAGPTLLPLGKLFWKSVKEDTLSTGABENAVCESPVETVRL
EEVSSLQVDSLETQPEPTDVKKESEKPKTPLMAFLQMSVRSSEGI PRSEEN
VKDSSCOTNSVEKTPSPPEPPAGTAOKNKETSSKDKKSVDKKSATENSQKNGKQ
EYREAPPCVQPTVEANAMQTGDKTPKSEKRRQSLGGFLKGLGPKRMSDAQVQTDPV
SIGPVGKSK"
```

ORIGIN

Query Match		12.0%; Score 86.6; DB 10; Length 2916;
Best Local Similarity		60.7%; Pred. No. 3.9e-13;
Matches 156; Conservative		2; Mismatches 96; Indels 3; Gaps 1;
QY	332	AGGGGGGAGCAGGACCCACCTTTGTAAACATGTTTGGGGGAATGTTTGTTCAC 391
Db	1314	AGGCAGGGCCCACTTACTACCGCTGGGAAAGTTGTTTGGAAAGAGTCAGTTAAGGAGG 1373
QY	392	TTTTCTTATTACCTGSCAAAATAATCCAGTGTGTGAGTCACCAGTAGAGATTATAA 451
Db	1374	ATACACTTTCCACAGTGCAGAGGAGACCGGTGTGTGAGTCACCAGTAGAGACCGTAA 1433
QY	452	AGTCCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAAGGAGATGCTGCAC 511
Db	1434	GGCTTCAGGAAGTAGAATCCAGCTTACAACTGTGGATCTCAGTGA--AGAGACCCAGC 1490
QY	512	CTGAACCCACWGAAGCGAACTCAAAAGAGAGAAAGCAAAACCAAGAACTCTCTGATGR 571
Db	1491	CTGAACCCACAGACGTAAAAGTCAAAGAGAAAGCAAAACCCCGAAGACCCCTCTGATGG 1550
QY	572	CGTTTCTCAGACAAATG 588
Db	1551	CGTTTCTCAGACAAATG 1567

Search completed: May 5, 2005, 03:58:10  
Job time : 3543.14 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:02:41 ; Search time 500.471 Seconds  
(without alignment)  
8551.893 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGCTTACC.....TGAATGGAATATGATTCG 723

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:\*

1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	722.2	99.9	723	2	AAV04697 Homo sapi
2	167.6	23.2	349	2	AAV88411 EST clone
3	167.6	23.2	2813	5	AA581115 DNA encod
4	167.6	23.2	2813	6	AA38092 Human BCA
5	167.6	23.2	2813	12	ADJ75127 Marker Ge
6	167.6	23.2	2813	12	ADQ18993 Human sof
7	167.6	23.2	2813	13	ADR25076 Breast ca
8	167.6	23.2	3066	2	AAV09025 Homo sapi
9	167.6	23.2	3418	12	ADQ23347 Human sof
10	166	23.0	2020	8	ACC90605 Human CGD
11	88.6	12.3	383	8	ABX44866 Bovine ES
12	70.2	9.7	540	13	ADQ53268 Novel can
13	41.2	5.7	957	6	ABZ13040 Arabidops
14	41.2	5.7	1347	3	AAC50675 Arabidops
15	41.2	5.7	1352	3	AAC33342 Arabidops
16	41.2	5.7	1352	3	AAC50674 Arabidops
17	41.2	5.7	1356	3	AAC36193 Arabidops
18	41.2	5.7	1470	3	AAC50673 Arabidops
19	41.2	5.7	1474	3	AAC39251 Arabidops
20	39.4	5.4	103464	13	ABD33278 Murine ca

c

c	21	37.8	5.2	2000	8	ADA71938	Ada71938 Rice gene
c	22	37.2	5.1	17534	6	ABK40026	ABK40026 Human che
	23	37.2	5.1	83836	11	ACN45080	ACN45080 Mouse gen
	24	36.6	5.1	3352	5	ABV27806	ABV27806 Human pro
	25	36.6	5.1	3353	5	ABV25280	ABV25280 Human pro
	26	36.4	5.0	428	4	AAI88208	AAI88208 Human pol
	27	36.2	5.0	198522	11	ACN44010	ACN44010 Human gen
c	28	36	5.0	598	12	ACH67256	ACH67256 Human gen
	29	36	5.0	885	10	ADF68708	ADF68708 Human met
	30	36	5.0	885	10	ADF89281	ADF89281 P gene se
	31	36	5.0	885	12	ADJ97169	ADJ97169 P gene DN
	32	36	5.0	885	12	ADL07820	ADL07820 P gene fr
	33	36	5.0	885	12	ADM67640	ADM67640 Human met
	34	36	5.0	7734	4	ABL29804	ABL29804 Drosophil
	35	36	5.0	13215	10	ADF68721	ADF68721 Human met
	36	36	5.0	13215	10	ADF89222	ADF89222 Human met
	37	36	5.0	13215	12	ADJ96810	ADJ96810 Human met
	38	36	5.0	13215	12	ADL07833	ADL07833 HMPV cDNA
	39	36	5.0	13215	12	ADM67281	ADM67281 Human met
c	40	35.8	5.0	110000	12	ADN06353	ADN06353 Human FLA
c	41	35.8	5.0	110000	13	ADS94372	ADS94372 Human 5-1
	42	35.4	4.9	13021	4	AAO4666	AAO4666 Human rep
	43	35.4	4.9	13021	4	ABL97573	ABL97573 Human tes
c	44	35.4	4.9	26997	4	AA46747	AA46747 Tumour su
	45	35.2	4.9	552	5	ADL43717	ADL43717 Human ova

## ALIGNMENTS

## RESULT 1

AAV04697  
ID AAV04697 standard; cDNA; 723 BP.

XX  
AC AAV04697;

XX  
DT 21-JUL-1998 (first entry)

XX  
DE Homo sapiens 20q13 amplicon 1b11 transcript.

XX  
KW 20q13 amplicon; chromosome 20; tumour; detection;  
XX  
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;  
XX  
KW treatment; age-related macular degeneration; retinitis pigmentation;  
XX  
KW Leber's congenital amaurosis; ds.

XX  
OS Homo sapiens.

XX  
FN WO9802539-A1.

XX  
PD 22-JAN-1998.

XX  
PF 15-JUL-1997; 97WO-US012343.

XX  
PR 15-JUL-1996; 96US-00680395.

XX  
PR 16-OCT-1996; 96US-00731499.

XX  
PR 17-JAN-1997; 97US-00785532.

XX  
(REGC) UNIV CALIFORNIA.

XX  
PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;

XX  
WPI; 1998-110587/10.

XX  
New sequences from the 20q13 amplicon - used for detecting chromosomal

XX  
PT abnormalities, particularly tumours, and for developing products for

XX  
PT treating diseases.

XX  
Claim 1; Page 61; 91pp; English.

XX  
The sequence is that of a cDNA sequence 1b11, which was isolated from the

XX  
20q13 amplicon. Its expression shows high correlation with the copy

XX  
number of the amplicon. It can be used as a probe for the detection of

XX  
chromosomal abnormalities at 20q13. It and other sequences isolated from

CC the 20q13 amplicon are consistently amplified in primary tumours. These  
 CC sequences are useful as probes or as probe targets for monitoring the  
 CC relative copy number of corresponding sequences from a biological sample  
 CC such as tumour cells. The sequences can also be used in therapeutic  
 CC applications for modulating the expression of the endogenous gene or the  
 CC activity of the gene product. Examples of therapeutic approaches include  
 CC antisense inhibition of gene expression, gene therapy, and monoclonal  
 CC antibodies that specifically bind the gene products. The products can  
 CC also be used in the treatment of other diseases, e.g. age-related macular  
 CC degeneration, Leber's congenital amaurosis and retinitis pigmentation  
 XX  
 SQ Sequence 723 BP; 210 A; 138 C; 178 G; 195 T; 0 U; 2 Other;

Query Match 99.9%; Score 722.2; DB 2; Length 723;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-222;  
 Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAAGCTGTCAATGGTTACCGTCTTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60  
 Db 1 TGAAGCTGTCAATGGTTACCGTCTTAACGTTGGACTCTTAAGAAATGATTATTCCTGG 60  
 QY 61 TTCTAGACAGGCCAAATGTAATTCACCTACGTCAGATTAAGAGGTGGCTTACTAG 120  
 Db 61 TTCTAGACAGGCCAAATGTAATTCACCTACGTCAGATTAAGAGGTGGCTTACTAG 120  
 QY 121 ATTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAGGACCTCTTATCCGT 180  
 Db 121 ATTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAGGACCTCTTATCCGT 180  
 QY 181 TCTTCCCTTTGGGGAAGGGCTTTGGCACTTCCATGTCATGTCAGTTGGAGTGGAA 240  
 Db 181 TCTTCCCTTTGGGGAAGGGCTTTGGCACTTCCATGTCATGTCAGTTGGAGTGGAA 240  
 QY 241 ATTGGTGGCTGTACAACTAGCATTTCTTCCAGATGTCCTGTGTAGAAATGGTC 300  
 Db 241 ATTGGTGGCTGTACAACTAGCATTTCTTCCAGATGTCCTGTGTAGAAATGGTC 300  
 QY 301 ATAGATTCAAACTGTAGTACTATGTGACAGGGGGGAGGACCCCACTTTGTAA 360  
 Db 301 ATAGATTCAAACTGTAGTACTATGTGACAGGGGGGAGGACCCCACTTTGTAA 360  
 QY 361 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTCTTATACCTGGCAAAATAATCCAG 420  
 Db 361 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTCTTATACCTGGCAAAATAATCCAG 420  
 QY 421 GTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAAGGAGTGAATCAGCCTTACAA 480  
 Db 421 GTGGTGTGAGTCACAGTAGAGATTATAAGTCCAAAGGAGTGAATCAGCCTTACAA 480  
 QY 481 ACAGTGGACCTCAACGAAGAGATGTCGACCTGAACCCCAACGAAACTCAAAAGA 540  
 Db 481 ACAGTGGACCTCAACGAAGAGATGTCGACCTGAACCCCAACGAAACTCAAAAGA 540  
 QY 541 GAAGAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTA 600  
 Db 541 GAAGAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTA 600  
 QY 601 CTTCCAGTATAGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAAACAATGGGCTCATGCC 660  
 Db 601 CTTCCAGTATAGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAAACAATGGGCTCATGCC 660  
 QY 661 ACAGGTAGTAGGAGACATAATCTAGCTGGTGTGATGGAATCTGAATGGAATATGATGATT 720  
 Db 661 ACAGGTAGTAGGAGACATAATCTAGCTGGTGTGATGGAATCTGAATGGAATATGATGATT 720  
 QY 721 GCG 723  
 Db 721 GCG 723

RESULT 2  
 AAV88411  
 ID AAV88411 standard; cDNA; 349 BP.

XX AAV88411;  
 AC  
 XX  
 DT 12-FEB-1999 (first entry)  
 XX  
 DE EST clone EK480.  
 XX  
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9845437-A2.  
 XX  
 PD 15-OCT-1998.  
 XX  
 PF 10-APR-1998; 98WO-US006956.  
 XX  
 PR 10-APR-1997; 97US-00837312.  
 XX  
 PA (GEMY ) GENETICS INST INC.  
 XX  
 PI Jacobs K, McCoy JM, Lavallie ER, Racie LA, Merberg D, Treacy M;  
 PI Spaulding V, Agostino MJ;  
 XX  
 DR WPI; 1999-070078/06.  
 XX

New polynucleotides encoding human secreted proteins - derived from e.g.  
 human blood, kidney, foetal lung, placenta, testes, brain, ovary,  
 pituitary, retina and colon cDNA libraries.

Claim 1; Page 388; 641pp; English.

The present sequence represents an expressed sequence tag (EST), and is a  
 polynucleotide of the invention. The polynucleotides of the invention are  
 all secreted EST sequences isolated from a variety of human tissue  
 sources. The EST sequences and proteins encoded by them are predicted to  
 have useful biological activities which would make them suitable for  
 treating, preventing or ameliorating medical conditions in humans and  
 animals, although no supporting data is given. Suggested activities  
 include nutritional activity, immune stimulating or suppressing activity,  
 haematopoiesis regulating activity, tissue growth activity,  
 activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 activity. The EST sequences are also stated to be useful for gene therapy

Query Match 23.2%; Score 167.6; DB 2; Length 349;  
 Best Local Similarity 98.2%; Pred. No. 2.5e-43;  
 Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCACCAAGGAGATTAAGAGTCCAAAGGAGTAGAATCAGCCTTAC 478  
 Db 69 ATGTGTGTGTGAGTCACCAAGGAGATTAAGAGTCCAAAGGAGTAGAATCAGCCTTAC 128  
 QY 479 AAACAGTGACCTCAACGAAGGAGATGTCGACCTGAACCCACGAAAGCGAACTCAAAA 538  
 Db 129 AAACAGTGACCTCAACGAAGGAGATGTCGACCTGAACCCACGAAAGCGAACTCAAAA 188  
 QY 539 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 598  
 Db 189 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 238

RESULT 3  
 AAS81115  
 ID AAS81115 standard; cDNA; 2813 BP.  
 XX  
 AC AAS81115;

```

XX 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #16919.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX P-PSDB; ASG16928.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 1; SEQ ID NO 16919; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
XX coding sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
XX
XX Query Match 23.2%; Score 167.6; DB 5; Length 2813;
XX Best Local Similarity 98.2%; Pred. No. 7.6e-43;
XX Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
XX 419 AGGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 478
XX 1298 ATGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 1357
XX
XX 479 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGAAACCCACGAGCGAACTCAAAA 538
XX 1358 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGAAACCCACGAGCGAACTCAAAA 1417
XX
XX 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
XX 1418 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467

```

```

RESULT 4
AAD38092
XX AAD38092 standard; DNA; 2813 BP.
XX
XX AAD38092;
XX
XX 10-SEP-2002 (first entry)
XX
XX Human BCAS1 (breast cancer amplified sequence 1) DNA.
XX
XX Human; BCAS1; breast cancer amplified sequence 1; AIBCL1; inflammation;
XX amplified in breast cancer 1; NABCL1; novel amplified in breast cancer 1;
XX hyperproliferative disorder; breast; prostate; cancer; prophylaxis;
XX infection; antisense therapy; cytostatic; antinflammatory; tumour; gene;
XX ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 118..1872
XX FT /*tag= a
XX FT /product= "Human BCAS1 protein"
XX
XX WO200231136-A1.
XX
XX 18-APR-2002.
XX
XX 09-OCT-2001; 2001WO-US031484.
XX
XX 11-OCT-2000; 2000US-00689255.
XX
XX (ISIS-) ISIS PHARM INC.
XX
XX Cowsett LM, Freier SM;
XX
XX WPI; 2002-444179/47.
XX
XX P-PSDB; AAE23791.
XX
XX New antisense compounds targeted to a nucleic acid molecule encoding
XX BCAS1, useful for treating diseases or conditions associated with BCAS1,
XX such as hyperproliferative disease, particularly breast or prostate
XX cancer.
XX
XX Example 10; Page 92-95; 104pp; English.
XX
XX The invention relates to antisense compounds, compositions and methods
XX for modulating the expression of BCAS1 (breast cancer amplified sequence
XX 1, also known as AIBCL1 for amplified in breast cancer 1 and NABCL1 for
XX novel amplified in breast cancer 1). The antisense compounds of the
XX invention are useful for treating an animal having a disease or condition
XX associated with BCAS1, such as hyperproliferative disorders including
XX breast or prostate cancer. These compounds are also used as research
XX reagents and diagnostics; to distinguish between functions of various
XX members of a biological pathway; in the treatment of a disease or
XX disorder, which can be treated by modulating the expression of BCAS1; as
XX prophylaxis, e.g. to prevent or delay infection, inflammation or tumour
XX formation, and as probes or primers. These antisense compounds are used
XX in antisense therapy. The present sequence is human BCAS1 DNA
XX
XX Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
XX
XX Query Match 23.2%; Score 167.6; DB 6; Length 2813;
XX Best Local Similarity 98.2%; Pred. No. 7.6e-43;
XX Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
XX 419 AGGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 478
XX 1298 ATGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCAGGAAGTAGAATCAGCCTTAC 1357
XX
XX 479 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGAAACCCACGAGCGAACTCAAAA 538
XX 1358 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGAAACCCACGAGCGAACTCAAAA 1417
XX
XX 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588
XX 1418 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467

```

```
QY 539 GAGAGAAAGCAAAACCAAGAACTCTCTGATGCGTTTCTCAGACAAATG 588
|||||
Db 1418 GAGAGAAAGCAAAACCAAGAACTCTCTGATGCGTTTCTCAGACAAATG 1467
|||||

RESULT 5
ADJ75127
ID ADJ75127 standard; DNA; 2813 BP.
XX
AC ADJ75127;
XX
DT 20-MAY-2004 (first entry)
XX
DE Marker gene SEQ ID NO:379.
XX
KW bronchial asthma; chronic obstructive pulmonary disease;
KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
KW gene therapy; marker gene; gene; ds.
XX
OS Homo sapiens.
XX
PN EP1394274-A2.
XX
PD 03-MAR-2004.
XX
PF 04-AUG-2003; 2003EP-00254857.
XX
PR 06-AUG-2002; 2002JP-00229312.
PR 20-MAR-2003; 2003JP-00077212.
XX
PA (GENO-) GENOX RES INC.
XX
PI Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;
XX
DR WPI; 2004-193155/19.
XX
PT Testing for bronchial asthma or chronic obstructive pulmonary disease by
PT comparing the expression level of a marker gene in a biological sample
PT from a subject with the expression level of the gene in a sample from a
PT healthy subject.
XX
PS Claim 1; SEQ ID NO 379; 24lpp; English.
XX
CC The present invention describes a method of testing for bronchial asthma
CC or chronic obstructive pulmonary disease. The method comprises
CC determining the expression level of a marker gene in a biological sample
CC from a subject, comparing the expression level determined with the
CC expression level of the marker gene in a biological sample from a healthy
CC subject, and judging whether the subject has bronchial asthma or chronic
CC obstructive pulmonary disease. The marker gene comprises: (a) a group of
CC genes (S1) whose expression levels increase when respiratory epithelial
CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)
CC whose expression levels decrease when respiratory epithelial cells are
CC stimulated with interleukin-13. Also described: (1) a reagent (I) for
CC testing for bronchial asthma or chronic obstructive pulmonary disease;
CC (2) a kit for screening for a candidate compound for a therapeutic agent
CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
CC an animal model for bronchial asthma or chronic obstructive pulmonary
CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
CC method for producing an animal model for bronchial asthma or chronic
CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial
CC asthma or chronic obstructive pulmonary disease, comprising the compound,
CC a marker gene or an antisense nucleic acid corresponding to a portion of
CC the marker gene, a ribozyme, a polynucleotide that suppresses the
CC expression of the gene through an RNAi effect or an antibody recognising
CC a protein encoded by a marker gene; and (7) a DNA chip for testing for
CC bronchial asthma or a chronic obstructive pulmonary disease, on which a
CC probe has been immobilised to assay a marker gene. (I) has respiratory
CC and antiasthmatic activities, and can be used in gene therapy. The method
CC is useful for testing for or screening for a therapeutic agent for
CC bronchial asthma or chronic obstructive pulmonary disease. The present
CC sequence is used in the exemplification of the present invention.
```

```
XX
SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
Query Match 23.2%; Score 167.6; DB 12; Length 2813;
Best Local Similarity 98.2%; Pred. No. 7.6e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACAGTAGAGATTATAAATCCAAAGAGTAGAATCAGCCTTAC 478
1298 ATGTGGTGTGAGTCACAGTAGAGATTATAAATCCAAAGAGTAGAATCAGCCTTAC 1357
Db
QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACACWAGGAACTCAAAA 538
1358 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACACWAGGAACTCAAAA 1417
Db
QY 539 GAGAGAAAGCAAAACCAAGAACTCTCTGATGCGTTTCTCAGACAAATG 588
1418 GAGAGAAAGCAAAACCAAGAACTCTCTGATGCGTTTCTCAGACAAATG 1467
Db

RESULT 6
ADQ18993
ID ADQ18993 standard; DNA; 2813 BP.
XX
AC ADQ18993;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 1812.
DE soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
DR WPI; 2004-441208/41.
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 1812; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytostatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC DNA of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;
Query Match 23.2%; Score 167.6; DB 12; Length 2813;
```

Best Local Similarity 98.2%; Pred. No. 7.6e-43;  
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACCAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 478  
DB 1298 ATGTGGTGTGAGTCACCAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1357

QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 538  
DB 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1417

QY 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGGTTTCTCAGACAAATG 588  
DB 1418 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGGTTTCTCAGACAAATG 1467

RESULT 7  
ADR25076  
ID ADR25076 standard; DNA; 2813 BP.  
XX  
AC ADR25076;  
XX  
DT 21-OCT-2004 (first entry)  
XX  
DE Breast cancer prognosis marker #937.  
XX  
KW ds; breast cancer; prognosis; gene expression; diagnosis.  
XX  
OS Homo sapiens.  
XX  
FN WO2004065545-A2.  
XX  
PD 05-AUG-2004.  
XX  
PF 15-JAN-2004; 2004WO-US001100.  
XX  
PR 15-JAN-2003; 2003US-00342887.  
XX  
PA (ROSE-) ROSETTA INPHARMATICS LLC.  
PA (NECA-) NETHERLANDS CANCER INST.  
XX  
PI Van't Veer LJ, He Y;  
XX  
DR WPI; 2004-593473/57.  
XX  
PT Classifying a breast cancer patient according to prognosis comprises  
PT determining the similarity between the level of expression of each of  
PT five genes in a cell sample taken from patient, to control levels.  
XX  
PS Disclosure; SEQ ID NO 937; 226pp; English.  
XX  
CC The invention relates to a method of classifying a breast cancer patient  
CC according to prognosis by determining the similarity between the level of  
CC expression of each of five genes for which markers are listed in the  
CC specification, in a cell sample taken from the breast cancer patient, to  
CC control levels of expression for each respective five genes to obtain a  
CC patient similarity value. The methods are useful for classifying a breast  
CC cancer patient according to prognosis. Kits and computer program products  
CC are useful for data analysis using the diagnostic, prognostic and  
CC statistical methods of the invention. This sequence corresponds to a  
CC marker used in the method of the invention.  
XX  
SQ Sequence 2813 BP; 868 A; 671 C; 758 G; 516 T; 0 U; 0 Other;

Query Match 23.2%; Score 167.6; DB 13; Length 2813;  
Best Local Similarity 98.2%; Pred. No. 7.6e-43;  
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACCAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 478  
DB 1298 ATGTGGTGTGAGTCACCAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1357

QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 538

DB 1358 AAACAGTGGACCTCAACGAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1417  
QY 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGGTTTCTCAGACAAATG 588  
DB 1418 GAGAAGAAAGCAACCAAGAACCTCTCTGATGCGGTTTCTCAGACAAATG 1467

RESULT 8  
AAV09025  
ID AAV09025 standard; cDNA; 3066 BP.  
XX  
AC AAV09025;  
XX  
DT 21-JUL-1998 (first entry)  
XX  
DE Homo sapiens 20q13 amplicon 1b1 transcript.  
XX  
KW 20q13 amplicon; chromosome 20; tumour; detection;  
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;  
KW treatment; age-related macular degeneration; retinitis pigmentata;  
KW Leber's congenital amaurosis; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO9802539-A1.  
XX  
PD 22-JAN-1998.  
XX  
PF 15-JUL-1997; 97WO-US012343.  
XX  
PR 15-JUL-1996; 96US-00680395.  
PR 16-OCT-1996; 96US-00731499.  
PR 17-JAN-1997; 97US-00785532.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;  
XX  
DR WPI; 1998-110587/10.  
XX  
PT New sequences from the 20q13 amplicon - used for detecting chromosomal  
PT abnormalities, particularly tumours, and for developing products for  
PT treating diseases.  
XX  
PS Claim 1; Page 69-70; 91pp; English.  
XX  
CC The sequence is that of a cDNA sequence 1b1, which was isolated from the  
CC 20q13 amplicon, it is overexpressed in numerous breast cancer cell lines  
CC and some primary tumours. It can be used as a probe for the detection of  
CC chromosomal abnormalities at 20q13. It and other sequences isolated from  
CC the 20q13 amplicon are consistently amplified in primary tumours. These  
CC sequences are useful as probes or as probe targets for monitoring the  
CC relative copy number of corresponding sequences from a biological sample  
CC such as tumour cells. The sequences can also be used in therapeutic  
CC applications for modulating the expression of the endogenous gene or the  
CC activity of the gene product. Examples of therapeutic approaches include  
CC antisense inhibition of gene expression, gene therapy, and monoclonal  
CC antibodies that specifically bind the gene products. The products can  
CC also be used in the treatment of other diseases, e.g. age-related macular  
CC degeneration, Leber's congenital amaurosis and retinitis pigmentation  
XX  
SQ Sequence 3066 BP; 944 A; 741 C; 818 G; 563 T; 0 U; 0 Other;

Query Match 23.2%; Score 167.6; DB 2; Length 3066;  
Best Local Similarity 98.2%; Pred. No. 7.9e-43;  
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGGTGGTGTGAGTCACCAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 478  
DB 1433 ATGTGGTGTGAGTCACCAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 1492

QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGACCTGAAACCCACGAAAGCGAACTCAAAA 538

```
|||||
Db 1493 AAACAGTGGACCTAACGGAAGGAGATGCTGCACCTGAACCCACGAGGCGAACTCAAAA 1552
QY 539 GAGAGAAAGCAAAACCAAGAACTCTCTCATGRCGTTTCTCAGACAAATG 588
Db 1553 GAGAGAAAGCAAAACCAAGAACTCTCTCATGRCGTTTCTCAGACAAATG 1602

RESULT 9
ADQ23347
ID ADQ23347 standard; DNA; 3418 BP.
XX AC
XX ADQ23347;
XX DT
XX 26-AUG-2004 (first entry)
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 6167.
XX
XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
KW ds.
XX Homo sapiens.
XX OS
XX WO2004048938-A2.
XX FN
XX 10-JUN-2004.
XX PD
XX 26-NOV-2003; 2003WO-US038193.
XX PF
XX 26-NOV-2002; 2002US-0429739P.
XX PR
XX (PROT-) PROTEIN DESIGN LABS INC.
XX PA
XX Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX DR
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
XX Example 2; SEQ ID NO 6167; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytostatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 3418 BP; 1029 A; 763 C; 842 G; 669 T; 0 U; 115 Other;
XX

Query Match 23.2%; Score 167.6; DB 12; Length 3418;
Best Local Similarity 98.2%; Pred. No. 8.4e-43;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTTGTGTGAGTCACAGTAGAGATTATTAAGTCCAGGAGTAGAATCGCCTTAC 478
Db 1346 ATGTGTTGTGTGAGTCACAGTAGAGATTATTAAGTCCAGGAGTAGAATCGCCTTAC 1405
QY 479 ARAACGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCACGAGGCGAACTCAAAA 538
Db 1406 ARAACGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCACGAGGCGAACTCAAAA 1465
QY 539 GAGAGAAAGCAAAACCAAGAACTCTCTCATGRCGTTTCTCAGACAAATG 588
```

```
|||||
Db 1466 GAGAGAAAGCAAAACCAAGAACTCTCTCATGRCGTTTCTCAGACAAATG 1515

RESULT 10
ACC90605
ID ACC90605 standard; DNA; 2020 BP.
XX AC
XX ACC90605;
XX DT
XX 12-AUG-2003 (first entry)
XX DE
XX Human CGDD-28 encoding DNA.
XX
XX Human; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; anti-allergic;
KW antiinflammatory; gynaecological; cancer; atherosclerosis; epilepsy;
KW Huntington's disease; stroke; AIDS; allergy; placenta; reproductive;
KW CGDD; cell growth; cell differentiation; cell death; gene; ds.
XX
XX Homo sapiens.
XX OS
XX WO2003027263-A2.
XX FN
XX 03-APR-2003.
XX PD
XX 26-SEP-2002; 2002WO-US031095.
XX PF
XX 28-SEP-2001; 2001US-0326389P.
XX PR
XX 05-OCT-2001; 2001US-0327380P.
XX PR
XX 05-OCT-2001; 2001US-0328186P.
XX PR
XX 12-OCT-2001; 2001US-0329690P.
XX PR
XX 26-OCT-2001; 2001US-0345384P.
XX PR
XX 26-OCT-2001; 2001US-0348165P.
XX PR
XX 02-NOV-2001; 2001US-0350219P.
XX PR
XX 09-NOV-2001; 2001US-0344518P.
XX PR
XX 09-NOV-2001; 2001US-0345143P.
XX PR
XX 16-NOV-2001; 2001US-0332375P.
XX PR
XX 03-DEC-2001; 2001US-0336908P.
XX PR
XX 07-DEC-2001; 2001US-0340747P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX PA
XX Azimzai Y, Baughn MR, Borowsky ML, Chawla NK;
XX Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ, Gorvad AE;
XX Griffin JA, Hafalia AJA, Ison CH, Kable AE, Kalafus DP;
XX Lehr-Mason PM, Lu DAM, Marquis JP, Nguyen DB, Ramkumar J;
XX Richardson IW, Sapperstein SK, Swarnakar A, Tang YT, Tran UK;
XX Warren BA, Xu Y, Yao MG, Yue H, Yue H;
XX WPI; 2003-421159/39.
XX DR
XX P-PSDB; ABR69628.
XX
XX New human proteins associated with cell growth, differentiation, and
XX death (CGDD), useful for diagnosing, treating and preventing diseases or
XX conditions associated with the aberrant CGDD expression e.g. cancer,
XX AIDS, or epilepsy.
XX
XX Claim 5; Page 330; 350pp; English.
XX
XX The invention relates to an isolated polypeptide associated with cell
XX growth, differentiation and death (CGDD). Also disclosed are the
XX polynucleotides encoding the polypeptides. The polypeptides and
XX polynucleotides are useful in diagnosing, treating and preventing
XX diseases or conditions associated with the decreased expression or over
XX expression of CGDD. Such diseases include cell proliferative (e.g.
XX cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
XX disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
XX reproductive disorders, or disorders of the placenta. They are also
XX useful in assessing the effects of exogenous compounds on the expression
XX of nucleic acid and amino acid sequences of CGDD. The CGDD or its
XX fragments are useful in screening compounds for effectiveness as an
XX agonist or antagonist of the polypeptides, or in altering the expression
```

CC of the target polynucleotide and compounds that specifically bind to or  
 CC modulate the activity of the polypeptide. Microarrays consisting  
 CC polynucleotides of the invention are useful in monitoring or measuring  
 CC protein-protein interactions, drug-target interactions, and gene  
 CC expression profiles. Sequences given in records ACC90578-ACC90634  
 CC represent polynucleotides encoding CGDD proteins of the invention  
 XX  
 SQ Sequence 2020 BP; 621 A; 525 C; 529 G; 345 T; 0 U; 0 Other;  
 Query Match 23.0%; Score 166; DB 8; Length 2020;  
 Best Local Similarity 97.6%; Pred. No. 2.1e-42;  
 Matches 166; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 QY 419 AGTGCTGTGAGTCACCACTAGAGATTATAAGTCAAGAGTAGAATCAGCCTTAC 478  
 DB 1312 ATGTGCTGTGAGTCACCACTAGAGATTATAAGTCAAGAGTAGAATCAGCCTTAC 1371  
 QY 479 AAACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAAACCCACGACGAACTCAAAA 538  
 DB 1372 AAACAGTGGACCTCAACGAAGGAGATGCTGCCTGAAACCCACGACGAACTCAAAA 1431  
 QY 539 GAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588  
 DB 1432 GAGAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 1481  
 RESULT 11  
 ID ABX44866 standard; cDNA; 383 BP.  
 XX  
 AC ABX44866;  
 XX  
 DT 21-FEB-2003 (first entry)  
 XX  
 DE Bovine EST associated with lactation/muscle/fat deposition #10031.  
 XX  
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
 KW muscle deposition; fat deposition; genome mapping; gene identification;  
 KW gene analysis; cattle breeding.  
 XX  
 OS Bos Taurus.  
 XX  
 PN US2002137139-A1.  
 XX  
 PD 26-SEP-2002.  
 XX  
 PF 24-SEP-2001; 2001US-00960352.  
 XX  
 PR 12-JAN-1999; 99US-0115707P.  
 PR 11-JAN-2000; 2000US-00480902.  
 XX  
 PA (BYAT/) BYATT J C.  
 PA (MATH/) MATHIALAGAN N.  
 PA (TAON/) TAO N.  
 PA (WARR/) WARREN W C.  
 XX  
 PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
 XX  
 DR WPI; 2003-110599/10.  
 XX  
 PT New nucleic acid associated with lactation, and muscle and fat  
 PT deposition, useful for genome mapping, gene identification and analysis,  
 PT cattle breeding, or for genetically improving cattle.  
 XX  
 PS Claim 2; SEQ ID NO 10031; 245pp; English.  
 XX  
 CC The invention relates to a purified nucleic acid molecule associated with  
 CC lactation or muscle and fat deposition (designated LMFD), derived from  
 CC cattle, and the LMFD nucleic acid can specifically hybridise to a second  
 CC nucleic acid molecule comprising any of 15112 nucleotide sequences,  
 CC appearing as ABX34836-ABX49947, or complements of them. Also included are  
 CC ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic  
 CC acid linked to a promoter and a 3' non- translated sequence that

CC functions in the cell to cause termination of transcription and addition  
 CC of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and  
 CC (2) determining a level or pattern of a molecule in a bovine cell or  
 CC tissue comprising: (a) incubating a marker nucleic acid (comprising any  
 CC of the 15112 nucleic acid sequences or its complement or fragment) with a  
 CC complementary nucleic acid molecule obtained from the bovine cell or  
 CC tissue, where hybridisation between the marker nucleic acid and the  
 CC complementary nucleic acid permits the detection of the molecule; and (b)  
 CC detecting the level or pattern of the complementary nucleic acid, where  
 CC the detection of the complementary nucleic acid is predictive of the  
 CC level or pattern of the molecule. The LMFD nucleic acid is used for  
 CC determining a level or pattern of a molecule in a bovine cell or tissue.  
 CC It is useful for genome mapping, gene identification and analysis, cattle  
 CC breeding, preparation of constructs for use in cattle gene expression, or  
 CC for genetically improving cattle. The present sequence is one of the  
 CC 15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The  
 CC present sequence was not shown in the specification but was obtained in  
 CC electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?DocID=20020137139  
 XX  
 SQ Sequence 383 BP; 123 A; 95 C; 108 G; 56 T; 0 U; 1 Other;  
 Query Match 12.3%; Score 88.6; DB 8; Length 383;  
 Best Local Similarity 62.9%; Pred. No. 9.8e-18;  
 Matches 168; Conservative 2; Mismatches 91; Indels 6; Gaps 2;  
 QY 328 GGACAGGGGGGCGACGAAGGACCCCACTTTGTAAACATGTTTTGGGGAAATGTTTTGTTT 387  
 DB 51 GAAAAAGGGGGCCCCAGCTCTCTGCCCTCTGGCAAACTGTTCTGAAAAAGTCAGTTAAA 110  
 QY 388 TTCATTTTCTATTACCTGGCAAAATATCCAGTGTGTGTGATCACCAGTAGAGATT 447  
 DB 111 GAGGATTTCAGTCCCCACAGGTGCAGAGGAGAAATGTTGTGTGTGATCACCAGTAGAGCT 170  
 QY 448 ATAAAGTCCGAAGGAGTAGAATCAGCCTTACAAACAGTGGACCT---CAACGAAGGAGAT 504  
 DB 171 GTAAGTCTGAAGAGTAGAGTACAGCTTACAAACAGTGGATCTCGGCGAAGAGGAGAG 230  
 QY 505 GCTGCACCTGAACCCACGACGAGCGAAACTCAAAAAGAGAAGAAAGCAAA---CCAAGAAC 561  
 DB 231 CGCACACCCGATCCCGCAGAAGAGCACCACGAGAGAGAAGAAACGACAGCCCGAGGCC 290  
 QY 562 TCTCTGATGRCGTTTCTCAGACAAATG 588  
 DB 291 TCCCTGATGGCGTCTCTCAGACAGATG 317  
 RESULT 12  
 ID ADQ53268 standard; DNA; 540 BP.  
 XX  
 AC ADQ53268;  
 XX  
 DT 21-OCT-2004 (first entry)  
 XX  
 DE Novel canine microarray-related DNA sequence SeqID4570.  
 XX  
 KW canine microarray; drug screening; toxicity assay;  
 KW environmental pollutant; cellular response; gene expression profile;  
 KW toxic response; liver necrosis; fatty liver disease;  
 KW protein adduct formation; hepatitis; dog; ds.  
 XX  
 OS Canis familiaris.  
 XX  
 PN WO2004063324-A2.  
 XX  
 PD 29-JUL-2004.  
 XX  
 PF 05-MAY-2003; 2003WO-US013853.  
 XX  
 PR 03-MAY-2002; 2002US-0377240P.  
 XX  
 PA (GENE-) GENE LOGIC INC.

PA (PFIZ ) PFIZER PROD INC.  
 XX Diggins JC, Porter M, Wei T;  
 XX WPI; 2004-561890/54.  
 DR  
 XX New isolated nucleic acid molecule, useful for drug screening and  
 PT toxicity assays or for assessing the impact, including toxicity, of a  
 PT compound, pharmaceutical agent or environmental pollutant on a cell or  
 PT living organism.  
 XX  
 PS Claim 1; SEQ ID NO 4570; 41pp; English.  
 XX  
 CC This invention is related to a novel isolated canine nucleic acid  
 CC sequences and the construction of canine microarrays containing a  
 CC significant portion of the canine genome. The isolated canine nucleic  
 CC acid sequences of the invention may be useful for drug screening and  
 CC toxicity assays. The invention is therefore useful for assessing the  
 CC impact, including toxicity, of a compound, pharmaceutical agent or  
 CC environmental pollutant on a cell or living organism. The methods are  
 CC useful for detecting genes that are up- or down-regulated in canines in a  
 CC disease state. The sequences are useful as diagnostic agents or markers  
 CC to detect a cellular response in a sample individually or as part of a  
 CC gene expression profile. It is also useful as a target for agents that  
 CC modulate gene expression or activity. The database is useful for  
 CC producing electronic Northern blots that allow the user to determine the cell  
 CC type or tissue in which a given gene is expressed and to allow  
 CC determination of the abundance or expression level of a given gene in a  
 CC particular tissue or cell. The methods are useful for determining the  
 CC similarity of a toxic response to one or more individual compounds. The  
 CC methods are useful for predicting at least one toxic response or the  
 CC likelihood that a compound or test agent will induce various specific  
 CC pathologies such as those of the liver (liver necrosis, fatty liver  
 CC disease, protein adduct formation or hepatitis), those of the kidney,  
 CC heart, brain or testes, or other pathologies associated with at least one  
 CC of the toxins. The methods are also useful for predicting or elucidating  
 CC the potential cellular pathways influenced, induced or modulated by the  
 CC compound or test agent due to the similarity of the expression profile  
 CC compared to the profile induced by a known toxin. The present sequence is  
 CC that of a canine DNA sequence which was claimed for use during the  
 CC production of a canine microarray of the invention.  
 XX  
 SQ Sequence 540 BP; 167 A; 137 C; 140 G; 74 T; 0 U; 22 Other;

Query Match 9.7%; Score 70.2; DB 13; Length 540;  
 Best Local Similarity 80.2%; Pred. No. 1.1e-11;  
 Matches 101; Conservative 2; Mismatches 17; Indels 6; Gaps 2;  
 QY 469 TCAGCCTTACAAACAGTGGACCTCA---ACGAGGAGATGCTGCACCTGAACCCACWGAA 525  
 DB 1 TCAGNNNTACAAACAGTNGATCTCAATGAAGAGGAGATGCCACANNTGAACCCACAGAN 60  
 QY 526 GCGAACTCAAAGAGAAGAACCAACCA---AGAACCTCTCTGATGCGTTTCTCAGA 582  
 DB 61 GTAAAACTCAAAGAGAAGAAATACAAACCCAGGAAACATCCCTGATGCGTTTCTCAGA 120  
 QY 583 CAAATG 588  
 DB 121 CAAATG 126

RESULT 13  
 AB213040  
 ID AB213040 standard; DNA; 957 BP.  
 XX  
 AC AB213040;  
 XX  
 DT 21-JAN-2003 (first entry)  
 XX  
 DE Arabidopsis thaliana stress regulated gene SEQ ID NO 845.  
 XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.  
 XX

OS Arabidopsis thaliana.  
 XX  
 PN WO200216655-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 24-AUG-2001; 2001WO-US026685.  
 XX  
 PR 24-AUG-2000; 2000US-0227866P.  
 PR 26-JAN-2001; 2001US-0264647P.  
 PR 22-JUN-2001; 2001US-0300111P.  
 XX  
 PA (SCRI ) SCRIPES RES INST.  
 PA (SYGN ) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Harper JF, Kreps J, Wang X, Zhu T;  
 XX  
 DR WPI; 2002-304127/34.  
 XX  
 PT Identifying a stress condition to which a plant cell has been exposed and  
 PT producing plants with increased tolerance to these abiotic stresses.  
 XX  
 PS Claim 144; SEQ ID NO 845; 577pp + Sequence Listing; English.  
 XX  
 CC The invention relates to identifying a stress condition to which a plant  
 CC cell has been exposed, comprising: (a) contacting nucleic acid  
 CC representative of expressed polynucleotides in the plant cell with an  
 CC array or probes representative of the plant cell genome; and (b)  
 CC detecting a profile of expressed polynucleotides in the plant cell  
 CC characteristic of a stress response. The method is useful in the  
 CC production of transgenic plants, cells and seeds and in producing plants  
 CC with increased tolerance to abiotic stress. The present sequence is that  
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used  
 CC in methods of the invention. Note: The sequence data for this patent is  
 CC not represented in the printed specification but is based on sequence  
 CC information supplied to Derwent by the European Patent Office  
 XX

Sequence 957 BP; 259 A; 188 C; 237 G; 273 T; 0 U; 0 Other;  
 Query Match 5.7%; Score 41.2; DB 6; Length 957;  
 Best Local Similarity 47.2%; Pred. No. 0.035;  
 Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;  
 QY 382 TTGTTTTCATTTCTTTTACCTGGCAAAATAATCCAGGTGGTGTGAGTCACCACTA 441  
 DB 358 TGGATGTTTAGAGTCTTTGGACATGACAAAGATGGGTACTTTGATGGAGGCTTGCAGAA 417  
 QY 442 GAGATTATTAAGTCAAGGAAGTAGAATCAGCCTTACAAACAGTGGACCTCAACGAAGA 501  
 DB 418 TGGCGTGTCTTCAGGCTATGACGTTGAATCCAGTGTTCAAATGATGCCATTTTGAAGGC 477  
 QY 502 GATGCTGCACCTGAACCCACWGAAGCGAAACTCAAAGAGAAGAAAGCAACCAAGAACC 561  
 DB 478 AGTGACGCACTGAGGCTATAGAGAAGATCTATCAAGGCAACAATTTAGCCCAATACT 537  
 QY 562 TCTCTGATGRCGTCTTTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGAAACCTAA 621  
 DB 538 TTTTCAGACGAAGTTCGGGCCACATCTTGTATTGGCACTTGTATGAGTGAAGGAACATC 597  
 QY 622 GATACCTAGA 631  
 DB 598 GAGGATAAGA 607

RESULT 14  
 AAC50675  
 ID AAC50675 standard; DNA; 1347 BP.  
 XX  
 AC AAC50675;  
 XX  
 DT 18-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65724.



XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 23-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 29-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 23-APR-1999; 99US-0130891P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 30-APR-1999; 99US-0132407P.

PR 04-MAY-1999; 99US-0132484P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 06-MAY-1999; 99US-0132487P.

PR 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 18-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 20-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137503P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 16-JUN-1999; 99US-0139453P.

PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145242P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.

PR	31-AUG-1999;	99US-0151438P.	RESULT 15	AC	AAC33342;	Arabidopsis thaliana DNA fragment SEQ ID NO: 2682.
PR	01-SEP-1999;	99US-0151930P.	AAC33342	XX		Hybridisation assay; genetic mapping; gene expression control;
PR	07-SEP-1999;	99US-0152363P.	ID AAC33342 standard; DNA; 1352 BP.	XX		protein identification; signal transduction pathway; metabolic pathway;
PR	10-SEP-1999;	99US-0153070P.	XX	AC	AAC33342;	promoter; termination sequence; ss.
PR	13-SEP-1999;	99US-0153758P.	XX	DT	17-OCT-2000 (first entry)	
PR	15-SEP-1999;	99US-0154018P.	XX	XX		Arabidopsis thaliana
PR	16-SEP-1999;	99US-0154039P.	XX	DE		Arabidopsis thaliana
PR	20-SEP-1999;	99US-0154779P.	XX	KW		Hybridisation assay; genetic mapping; gene expression control;
PR	22-SEP-1999;	99US-0155139P.	XX	KW		protein identification; signal transduction pathway; metabolic pathway;
PR	24-SEP-1999;	99US-0155659P.	XX	OS		promoter; termination sequence; ss.
PR	28-SEP-1999;	99US-0156458P.	XX	XX		Arabidopsis thaliana
PR	29-SEP-1999;	99US-0156596P.	XX	PN	EP1033405-A2.	
PR	04-OCT-1999;	99US-0157117P.	XX	PD	06-SEP-2000.	
PR	05-OCT-1999;	99US-0157753P.	XX	XX		25-FEB-2000; 2000EP-00301439.
PR	06-OCT-1999;	99US-0157865P.	XX	XX		25-FEB-1999; 99US-0121825P.
PR	07-OCT-1999;	99US-0158029P.	XX	XX		99US-0123180P.
PR	08-OCT-1999;	99US-0158232P.	XX	XX		99US-0123548P.
PR	12-OCT-1999;	99US-0158369P.	XX	XX		99US-0125788P.
PR	13-OCT-1999;	99US-0159293P.	XX	XX		99US-0126264P.
PR	13-OCT-1999;	99US-0159294P.	XX	XX		99US-0126785P.
PR	13-OCT-1999;	99US-0159295P.	XX	XX		99US-0127462P.
PR	14-OCT-1999;	99US-0159329P.	XX	XX		99US-0128234P.
PR	14-OCT-1999;	99US-0159330P.	XX	XX		99US-0128714P.
PR	14-OCT-1999;	99US-0159331P.	XX	XX		99US-0129845P.
PR	14-OCT-1999;	99US-0159637P.	XX	XX		99US-0130077P.
PR	14-OCT-1999;	99US-0159638P.	XX	XX		99US-0130449P.
PR	18-OCT-1999;	99US-0159584P.	XX	XX		99US-0130510P.
PR	21-OCT-1999;	99US-0160741P.	XX	XX		99US-0130891P.
PR	21-OCT-1999;	99US-0160767P.	XX	XX		99US-0131449P.
PR	21-OCT-1999;	99US-0160768P.	XX	XX		99US-0132048P.
PR	21-OCT-1999;	99US-0160770P.	XX	XX		99US-0132407P.
PR	21-OCT-1999;	99US-0160814P.	XX	XX		99US-0132484P.
PR	21-OCT-1999;	99US-0160815P.	XX	XX		99US-0132485P.
PR	22-OCT-1999;	99US-0160880P.	XX	XX		99US-0132486P.
PR	22-OCT-1999;	99US-0160981P.	XX	XX		99US-0132487P.
PR	22-OCT-1999;	99US-0160989P.	XX	XX		99US-0132863P.
PR	25-OCT-1999;	99US-0161404P.	XX	XX		99US-0134256P.
PR	25-OCT-1999;	99US-0161405P.	XX	XX		99US-0134218P.
PR	25-OCT-1999;	99US-0161406P.	XX	XX		99US-0134219P.
PR	26-OCT-1999;	99US-0161359P.	XX	XX		99US-0134221P.
PR	26-OCT-1999;	99US-0161360P.	XX	XX		99US-0134370P.
PR	26-OCT-1999;	99US-0161361P.	XX	XX		99US-0134768P.
PR	28-OCT-1999;	99US-0161920P.	XX	XX		99US-0134941P.
PR	28-OCT-1999;	99US-0161922P.	XX	XX		99US-0135124P.
PR	28-OCT-1999;	99US-0161932P.	XX	XX		99US-0135353P.
PR	29-OCT-1999;	99US-0161933P.	XX	XX		99US-0135629P.
PR	29-OCT-1999;	99US-0162142P.	XX	XX		99US-0136021P.
PR	29-OCT-1999;	99US-0162143P.	XX	XX		99US-0136392P.
PR	29-OCT-1999;	99US-0162144P.	XX	XX		99US-0136782P.
PR	29-OCT-1999;	99US-0162145P.	XX	XX		99US-0137222P.
PR	29-OCT-1999;	99US-0162146P.	XX	XX		99US-0137528P.
PR	29-OCT-1999;	99US-0162147P.	XX	XX		99US-0137502P.
PR	29-OCT-1999;	99US-0162148P.	XX	XX		99US-0137724P.
PR	29-OCT-1999;	99US-0162149P.	XX	XX		99US-0138094P.
PR	29-OCT-1999;	99US-0162150P.	XX	XX		99US-0138540P.
PR	29-OCT-1999;	99US-0162151P.	XX	XX		99US-0138847P.
PR	29-OCT-1999;	99US-0162152P.	XX	XX		99US-0139119P.
PR	29-OCT-1999;	99US-0162153P.	XX	XX		99US-0139452P.
PR	29-OCT-1999;	99US-0162154P.	XX	XX		99US-0139453P.
PR	29-OCT-1999;	99US-0162155P.	XX	XX		99US-0139492P.
PR	29-OCT-1999;	99US-0162156P.	XX	XX		99US-0139454P.
PR	29-OCT-1999;	99US-0162157P.	XX	XX		99US-0139455P.
PR	29-OCT-1999;	99US-0162158P.	XX	XX		99US-0139456P.
PR	29-OCT-1999;	99US-0162159P.	XX	XX		99US-0139457P.
PR	29-OCT-1999;	99US-0162160P.	XX	XX		99US-0139458P.
PR	29-OCT-1999;	99US-0162161P.	XX	XX		99US-0139459P.
PR	29-OCT-1999;	99US-0162162P.	XX	XX		99US-0139460P.
PR	29-OCT-1999;	99US-0162163P.	XX	XX		99US-0139461P.
PR	29-OCT-1999;	99US-0162164P.	XX	XX		99US-0139462P.
PR	29-OCT-1999;	99US-0162165P.	XX	XX		99US-0139463P.
PR	29-OCT-1999;	99US-0162166P.	XX	XX		99US-0139464P.
PR	29-OCT-1999;	99US-0162167P.	XX	XX		99US-0139465P.
PR	29-OCT-1999;	99US-0162168P.	XX	XX		99US-0139466P.
PR	29-OCT-1999;	99US-0162169P.	XX	XX		99US-0139467P.
PR	29-OCT-1999;	99US-0162170P.	XX	XX		99US-0139468P.
PR	29-OCT-1999;	99US-0162171P.	XX	XX		99US-0139469P.
PR	29-OCT-1999;	99US-0162172P.	XX	XX		99US-0139470P.
PR	29-OCT-1999;	99US-0162173P.	XX	XX		99US-0139471P.
PR	29-OCT-1999;	99US-0162174P.	XX	XX		99US-0139472P.
PR	29-OCT-1999;	99US-0162175P.	XX	XX		99US-0139473P.
PR	29-OCT-1999;	99US-0162176P.	XX	XX		99US-0139474P.
PR	29-OCT-1999;	99US-0162177P.	XX	XX		99US-0139475P.
PR	29-OCT-1999;	99US-0162178P.	XX	XX		99US-0139476P.
PR	29-OCT-1999;	99US-0162179P.	XX	XX		99US-0139480P.
PR	29-OCT-1999;	99US-0162180P.	XX	XX		99US-0139484P.
PR	29-OCT-1999;	99US-0162181P.	XX	XX		99US-0139488P.
PR	29-OCT-1999;	99US-0162182P.	XX	XX		99US-0139492P.
PR	29-OCT-1999;	99US-0162183P.	XX	XX		99US-0139496P.
PR	29-OCT-1999;	99US-0162184P.	XX	XX		99US-0139500P.
PR	29-OCT-1999;	99US-0162185P.	XX	XX		99US-0139504P.
PR	29-OCT-1999;	99US-0162186P.	XX	XX		99US-0139508P.
PR	29-OCT-1999;	99US-0162187P.	XX	XX		99US-0139512P.
PR	29-OCT-1999;	99US-0162188P.	XX	XX		99US-0139516P.
PR	29-OCT-1999;	99US-0162189P.	XX	XX		99US-0139520P.
PR	29-OCT-1999;	99US-0162190P.	XX	XX		99US-0139524P.
PR	29-OCT-1999;	99US-0162191P.	XX	XX		99US-0139528P.
PR	29-OCT-1999;	99US-0162192P.	XX	XX		99US-0139532P.
PR	29-OCT-1999;	99US-0162193P.	XX	XX		99US-0139536P.
PR	29-OCT-1999;	99US-0162194P.	XX	XX		99US-0139540P.
PR	29-OCT-1999;	99US-0162195P.	XX	XX		99US-0139544P.
PR	29-OCT-1999;	99US-0162196P.	XX	XX		99US-0139548P.
PR	29-OCT-1999;	99US-0162197P.	XX	XX		99US-0139552P.
PR	29-OCT-1999;	99US-0162198P.	XX	XX		99US-0139556P.
PR	29-OCT-1999;	99US-0162199P.	XX	XX		99US-0139560P.
PR	29-OCT-1999;	99US-0162200P.	XX	XX		99US-0139564P.
PR	29-OCT-1999;	99US-0162201P.	XX	XX		99US-0139568P.
PR	29-OCT-1999;	99US-0162202P.	XX	XX		99US-0139572P.
PR	29-OCT-1999;	99US-0162203P.	XX	XX		99US-0139576P.
PR	29-OCT-1999;	99US-0162204P.	XX	XX		99US-0139580P.
PR	29-OCT-1999;	99US-0162205P.	XX	XX		99US-0139584P.
PR	29-OCT-1999;	99US-0162206P.	XX	XX		99US-0139588P.
PR	29-OCT-1999;	99US-0162207P.	XX	XX		99US-0139592P.
PR	29-OCT-1999;	99US-0162208P.	XX	XX		99US-0139596P.
PR	29-OCT-1999;	99US-0162209P.	XX	XX		99US-0139600P.
PR	29-OCT-1999;	99US-0162210P.	XX	XX		99US-0139604P.
PR	29-OCT-1999;	99US-0162211P.	XX	XX		99US-0139608P.
PR	29-OCT-1999;	99US-0162212P.	XX	XX		99US-0139612P.
PR	29-OCT-1999;	99US-0162213P.	XX	XX		99US-0139616P.
PR	29-OCT-1999;	99US-0162214P.	XX	XX		99US-0139620P.
PR	29-OCT-1999;	99US-0162215P.	XX	XX		99US-0139624P.
PR	29-OCT-1999;	99US-0162216P.	XX	XX		99US-0139628P.
PR	29-OCT-1999;	99US-0162217P.	XX	XX		99US-0139632P.
PR	29-OCT-1999;	99US-0162218P.	XX	XX		99US-0139636P.
PR	29-OCT-1999;	99US-0162219P.	XX	XX		99US-0139640P.
PR	29-OCT-1999;	99US-0162220P.	XX	XX		99US-0139644P.
PR	29-OCT-1999;	99US-0162221P.	XX	XX		99US-0139648P.
PR	29-OCT-1999;	99US-0162222P.	XX	XX		99US-0139652P.
PR	29-OCT-1999;	99US-0162223P.	XX	XX		99US-0139656P.
PR	29-OCT-1999;	99US-0162224P.	XX	XX		99US-0139660P.
PR	29-OCT-1999;	99US-0162225P.	XX	XX		99US-0139664P.
PR	29-OCT-1999;	99US-0162226P.	XX	XX		99US-0139668P.
PR	29-OCT-1999;	99US-0162227P.	XX	XX		99US-0139672P.
PR	29-OCT-1999;	99US-0162228P.	XX	XX		99US-0139676P.
PR	29-OCT-1999;	99US-0162229P.	XX	XX		99US-0139680P.
PR	29-OCT-1999;	99US-0162230P.	XX	XX		99US-0139684P.
PR	29-OCT-1999;	99US-0162231P.	XX	XX		99US-0139688P.
PR	29-OCT-1999;	99US-0162232P.	XX	XX		99US-0139692P.
PR	29-OCT-1999;	99US-0162233P.	XX	XX		99US-0139696P.
PR	29-OCT-1999;	99US-0162234P.	XX	XX		99US-0139700P.
PR	29-OCT-1999;	99US-0162235P.	XX	XX		99US-0139704P.
PR	29-OCT-1999;	99US-0162236P.	XX	XX		99US-0139708P.
PR	29-OCT-1999;	99US-0162237P.	XX	XX		99US-0139712P.
PR	29-OCT-1999;	99US-0162238P.	XX	XX		99US-0139716P.
PR	29-OCT-1999;	99US-0162239P.	XX	XX		99US-0139720P.
PR	29-OCT-1999;	99US-0162240P.	XX	XX		99US-0139724P.
PR	29-OCT-1999;	99US-0162241P.	XX	XX		99US-0139728P.
PR	29-OCT-1999;	99US-0162242P.	XX	XX		99US-0139732P.
PR	29-OCT-1999;	99US-0162243P.	XX	XX		99US-0139736P.
PR	29-OCT-1999;	99US-0162244P.	XX	XX		99US-0139740P.
PR	29-OCT-1999;	99US-0162245P.	XX	XX		99US-0139744P.
PR	29-OCT-1999;	99US-0162246P.	XX	XX		99US-0139748P.
PR	29-OCT-1999;	99US-0162247P.	XX	XX		99US-0139752P.
PR	29-OCT-1999;	99US-0162248P.	XX	XX		99US-0139756P.
PR	29-OCT-1999;	99US-0162249P.	XX	XX		99US-0139760P.
PR	29-OCT-1999;	99US-0162250P.	XX	XX		99US-0139764P.
PR	29-OCT-1999;	99US-0162251P.	XX	XX		99US-0139768P.
PR	29-OCT-1999;	99US-0162252P.	XX	XX		99US-0139772P.
PR	29-OCT-1999;	99US-0162253P.	XX	XX		99US-0139776P.
PR	29-OCT-1999;	99US-0162254P.	XX	XX		99US-0139780P.
PR	29-OCT-1999;	99US-0162255P.	XX	XX		99US-0139784P.
PR	29-OCT-1999					

```
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145221P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145919P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.

PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 5.7%; Score 41.2; DB 3; Length 1352;
Best Local Similarity 47.2%; Pred. No. 0.042;
Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;

Qy 382 TTGTTTTTTCATTTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCAACAGTA 441
Db 596 TGGATGTTTAGAGTCTTTGGACATGACAAAGTATGGGTACTTGTATGGAGGCTTGCCGAA 655

Qy 442 GAGATTATAAGTCCAAGGAAGTAGATCAGCCTTACAAACAGTGGACCTCAACGAAGA 501
Db 656 TGGCGTCTTCAGGCTATGACGTTGAACTCCAGTGTTCATATGATGCAATTTGAAAGCC 715

Qy 502 GATGCTCACCTGAACCCACGAGGCGAAACTCAAAAGAGAGAAAGCAAAACCAAGAAC 561
Db 716 AGTCGACCACTGAGGCTATAGAAAGATCTATCAGGACAAACATTTAGCCCAATAACT 775
```

QY	562	TCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGGAAACCTAA	621
Db	776	TTTCAGACGAAGTTCGGCCACATCTTGTATTGGCACTTGATCAGGTGAAGGAAACATC	835
QY	622	GATACCTAGA	631
Db	836	GAGGATAGA	845

Search completed: May 5, 2005, 00:37:07  
Job time : 507.471 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:46:20 ; Search time 162.426 Seconds  
(without alignments)  
7283.499 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGGTACC.....TGAATGGATATGGATTGG 723

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	722.2	99.9	723	2	US-08-680-395-2
2	722.2	99.9	723	4	US-08-892-695-2
3	714.8	98.9	130563	4	US-09-943-016-12273
4	714.8	98.9	131379	4	US-09-949-016-16050
5	167.6	23.2	2813	3	US-09-689-255C-3
6	167.6	23.2	2813	4	US-09-949-016-531
7	167.6	23.2	3066	4	US-08-892-695-12
8	167.6	23.2	3629	4	US-09-949-016-4308
9	41.8	5.8	7218	1	US-08-232-463-14
10	36	5.0	1728	4	US-09-248-796A-3540
11	34.8	4.8	601	4	US-09-949-016-162024
12	34.8	4.8	601	4	US-09-949-016-204212
13	34.8	4.8	9668	4	US-09-949-016-4532
14	34.8	4.8	28374	4	US-09-949-016-17508
15	34.8	4.8	238815	4	US-09-949-016-16274
16	34.4	4.8	415	4	US-09-270-767-8451
17	34.4	4.8	415	4	US-09-270-767-73733
18	34.4	4.8	601	4	US-09-949-016-204213
19	34.4	4.8	68035	4	US-09-949-016-16219
20	34.2	4.7	832	4	US-09-621-976-2813
21	34.2	4.7	45323	4	US-09-949-016-16142
22	34	4.7	3183	4	US-09-248-796A-2521
23	33.8	4.7	74353	4	US-09-949-016-15336
24	33.8	4.7	597	4	US-09-270-767-5349
25	33.8	4.7	597	4	US-09-270-767-20631
26	33.4	4.6	601	4	US-09-943-016-81376
27	33.4	4.6	1596	4	US-09-248-796A-1457

28	33.4	4.6	121982	4	US-09-949-016-12085	Sequence 12085, A
29	33.4	4.6	121982	4	US-09-949-016-14105	Sequence 14105, A
c 30	33.2	4.6	1584	1	US-07-667-276A-1	Sequence 1, Appli
31	33.2	4.6	2709	4	US-10-087-402-5	Sequence 5, Appli
32	33.2	4.6	2746	2	US-08-576-165-3	Sequence 3, Appli
c 33	33	4.6	371	3	US-08-943-731-103	Sequence 103, App
34	33	4.6	534	4	US-09-387-286-26	Sequence 26, Appl
c 35	33	4.6	558	4	US-09-270-767-7078	Sequence 7078, Ap
c 36	33	4.6	558	4	US-09-270-767-22360	Sequence 22360, A
37	33	4.6	601	4	US-09-949-016-37900	Sequence 37900, A
38	33	4.6	601	4	US-09-949-016-176986	Sequence 176986,
39	33	4.6	1236	4	US-09-540-236-1364	Sequence 1364, Ap
c 40	33	4.6	1989	4	US-09-248-796A-1350	Sequence 1350, Ap
c 41	33	4.6	38682	3	US-08-943-731-2	Sequence 2, Appli
c 42	33	4.6	40315	4	US-09-949-016-11753	Sequence 11753, A
c 43	33	4.6	40649	4	US-09-949-016-14219	Sequence 14219, A
c 44	33	4.6	41062	4	US-09-949-016-12761	Sequence 12761, A
c 45	33	4.6	41062	4	US-09-949-016-16796	Sequence 16796, A

## ALIGNMENTS

RESULT 1  
US-08-680-395-2  
; Sequence 2, Application US/08680395  
; Patent No. 5892010  
; GENERAL INFORMATION:  
; APPLICANT: Gray, Joe W.  
; APPLICANT: Colling, Colin  
; APPLICANT: Hwang, Soo-in  
; APPLICANT: Godfrey, Tony  
; APPLICANT: Kowbel, David  
; APPLICANT: Rommens, Johanna  
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,395  
; FILING DATE: 15-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-0689000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 723 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..723  
; OTHER INFORMATION: /note= "cdna clone 1b11 of 3.5kb  
; OTHER INFORMATION: transcript"  
US-08-680-395-2

Query Match		99.9%;	Score 722.2;	DB 2;	Length 723;
Best Local Similarity		100.0%;	Pred. No. 3.1e-228;		
Matches 723;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
; EARLIER FILING DATE: 1997-01-17					
; EARLIER APPLICATION NUMBER: 08/731,499					
; EARLIER FILING DATE: 1996-10-16					
; EARLIER APPLICATION NUMBER: 08/680,395					
; EARLIER FILING DATE: 1996-07-15					
; NUMBER OF SEQ ID NOS: 59					
; SOFTWARE: PatentIn Ver. 2.0					
; SEQ ID NO 2					
; LENGTH: 723					
; TYPE: DNA					
; ORGANISM: Artificial Sequence					
; FEATURE:					
; OTHER INFORMATION: Description of Artificial Sequence:lb11					
US-08-892-695-2					
Query Match 99.9%; Score 722.2; DB 4; Length 723;					
Best Local Similarity 100.0%; Pred. No. 3.1e-228;					
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	TGGAAGCTGTGTCATGGTTACCGTCTCTAAAGTTGGACTCTTAAAGAAAATGATTATTCCTGG	60		
DB	1	TGGAAGCTGTGTCATGGTTACCGTCTCTAAAGTTGGACTCTTAAAGAAAATGATTATTCCTGG	60		
QY	61	TTTCTAGACAGCCCAATGTAATTCACCTACGTGGCAGATTAAAGAGTGGGCTTACTAG	120		
DB	61	TTTCTAGACAGCCCAATGTAATTCACCTACGTGGCAGATTAAAGAGTGGGCTTACTAG	120		
QY	121	ATTGATTTGGGTATTTGAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	180		
DB	121	ATTGATTTGGGTATTTGAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	180		
QY	181	TCTTCCCTTTGGGAAAGGGCTTTTGGCACTTCCATGTCGAATGGCAGTTGAGCTTGGAA	240		
DB	181	TCTTCCCTTTGGGAAAGGGCTTTTGGCACTTCCATGTCGAATGGCAGTTGAGCTTGGAA	240		
QY	241	ATTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	300		
DB	241	ATTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	300		
QY	301	ATAGATTTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGACCCCACTTTGTAA	360		
DB	301	ATAGATTTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGACCCCACTTTGTAA	360		
QY	361	AAATGTTTGGGGAAATGTTTGTGTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT	420		
DB	361	AAATGTTTGGGGAAATGTTTGTGTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT	420		
QY	421	GTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	480		
DB	421	GTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	480		
QY	481	ACAGTGGACCTCAACGAGGAGATGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT	540		
DB	481	ACAGTGGACCTCAACGAGGAGATGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT	540		
QY	541	GAAGAAAGCAAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	600		
DB	541	GAAGAAAGCAAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT	600		
QY	601	CTTCCAGTATAGGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAAACAATGGGCTCATGCC	660		
DB	601	CTTCCAGTATAGGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAAACAATGGGCTCATGCC	660		
QY	661	ACAGGTAGTAGGACATTAATGTTAGTGTGTTATGGAATGGAATGGAATGGAATGGAAT	720		
DB	661	ACAGGTAGTAGGACATTAATGTTAGTGTGTTATGGAATGGAATGGAATGGAATGGAAT	720		
QY	721	GC 723			
DB	721	GC 723			
RESULT 2					
US-08-892-695-2					
; Sequence 2, Application US/08892695A					
; Patent No. 6808878					
; GENERAL INFORMATION:					
; APPLICANT: Gray, Joe W					
; APPLICANT: Collins, Collin					
; APPLICANT: Hwang, Soo In					
; APPLICANT: Godfrey, Tony					
; APPLICANT: Kowel, David					
; APPLICANT: Rommens, Johanna					
; TITLE OF INVENTION: GENES FROM THE 20Q13 AMPLICON AND THEIR USES					
; FILE REFERENCE: 2500.124US3					
; CURRENT APPLICATION NUMBER: US/08/892,695A					
; CURRENT FILING DATE: 1997-07-15					

; EARLIER APPLICATION NUMBER: 08/785,532  
; EARLIER FILING DATE: 1997-01-17  
; EARLIER APPLICATION NUMBER: 08/731,499  
; EARLIER FILING DATE: 1996-10-16  
; EARLIER APPLICATION NUMBER: 08/680,395  
; EARLIER FILING DATE: 1996-07-15  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 723  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:lb11  
US-08-892-695-2

Query Match 99.9%; Score 722.2; DB 4; Length 723;  
Best Local Similarity 100.0%; Pred. No. 3.1e-228; Indels 0; Gaps 0;  
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TGGAAGCTGTCATGGTTACCGTCTCTAAAGTTGGACTCTTAAGAAAATGATTATTCCTGG 60  
DB 1 TGGAAGCTGTCATGGTTACCGTCTCTAAAGTTGGACTCTTAAGAAAATGATTATTCCTGG 60  
QY 61 TTTCTAGACAGCCCAAAATGTAATTCACCTACGTGGCAGATTAAAGAGTGGGCTTACTAG 120  
DB 61 TTTCTAGACAGCCCAAAATGTAATTCACCTACGTGGCAGATTAAAGAGTGGGCTTACTAG 120  
QY 121 ATTGATTTGGGTATTCAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 180  
DB 121 ATTGATTTGGGTATTCAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 180  
QY 181 TCTTCCCTTTGGGAAAGGGCTTTGGCACTTCCATGTCGAATGGCAGTTGAGCTTGGAA 240  
DB 181 TCTTCCCTTTGGGAAAGGGCTTTGGCACTTCCATGTCGAATGGCAGTTGAGCTTGGAA 240  
QY 241 ATTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 300  
DB 241 ATTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 300  
QY 301 ATAGATTTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGACCCCACTTTGTAA 360  
DB 301 ATAGATTTCAAACTGTAGCTACTATGTGACAGGGGGGAGCAAGACCCCACTTTGTAA 360  
QY 361 AACATGTTTGGGGAAATGTTTGTGTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT 420  
DB 361 AACATGTTTGGGGAAATGTTTGTGTTTCACTTTTCACTTTTCACTTTTCACTTTTCACTTTT 420  
QY 421 GTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 480  
DB 421 GTGGTGGCTGTACAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 480  
QY 481 ACAGTGGACCTCAACGAGGAGATGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 540  
DB 481 ACAGTGGACCTCAACGAGGAGATGCTGACCTGACCTGACCTGACCTGACCTGACCTGACCT 540  
QY 541 GAAGAAAGCAAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 600  
DB 541 GAAGAAAGCAAACTAGCATGCTCTGAATGACAGTCCCAAAAAGGACCTCTTATCCGT 600  
QY 601 CTTCCAGTATAGGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAAACAATGGGCTCATGCC 660  
DB 601 CTTCCAGTATAGGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAAACAATGGGCTCATGCC 660  
QY 661 ACAGGTAGTAGGACATTAATGTTAGTGTGTTATGGAATGGAATGGAATGGAATGGAAT 720  
DB 661 ACAGGTAGTAGGACATTAATGTTAGTGTGTTATGGAATGGAATGGAATGGAATGGAAT 720  
QY 721 GC 723  
DB 721 GC 723

RESULT 3  
US-09-949-016-12273  
; Sequence 12273, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12273  
; LENGTH: 130563  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12273

Query Match 98.9%; Score 714.8; DB 4; Length 130563;  
Best Local Similarity 99.2%; Pred. No. 1.8e-224;  
Matches 716; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 TGGAGCTGTCATGTTACCGTCTCTAAGCTTGGACTCTTAAGAAATGATTATTCCTGG 60  
DB 105047 TGGAGCTGTCATGTTACCGTCTCTAAGCTTGGACTCTTAAGAAATGATTATTCCTGG 105106  
  
QY 61 TTCTAGACAGCCAAATGTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 120  
DB 105107 TTCTAGACAGCCAAATGTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 105166  
  
QY 121 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 180  
DB 105167 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 105226  
  
QY 181 TCTTCCCTTGGGGAAGGCTTTTGGCACTTCCATGTCATATGTCAGTGGCCTTGAA 240  
DB 105227 TCTTCCCTTGGGGAAGGCTTTTGGCACTTCCATGTCATATGTCAGTGGCCTTGAA 105286  
  
QY 241 ATTTGATGCTGTACACATAGCAATTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 300  
DB 105287 ATTTGATGCTGTACACATAGCAATTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 105346  
  
QY 301 ATAGATTCAAACCTAGTACTATGTGGACAGGGGGGCGACAGGACCCCACTTTGTAA 360  
DB 105347 ATAGATTCAAACCTAGTACTATGTGGACAGGGGGGCGACAGGACCCCACTTTGTAA 105406  
  
QY 361 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTTCTTATTTACCTGGCAAAATATCCAG 420  
DB 105407 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTTCTTATTTACCTGGCAAAATATCCAG 105466  
  
QY 421 GTGGTGTGTAGTCCAGTAGAGATTATAAGTCCAGAGGAGTAGAATCAGCCTTACAA 480  
DB 105467 GTGGTGTGTAGTCCAGTAGAGATTATAAGTCCAGAGGAGTAGAATCAGCCTTACAA 105526  
  
QY 481 ACAGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCCAAGGAGCGAACTCAAAAGA 540  
DB 105527 ACAGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCCAAGGAGCGAACTCAAAAGA 105586  
  
QY 541 GAAGAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATGTAAGCCCTTA 600  
DB 105587 GAAGAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATGTAAGCCCTTA 105646  
  
QY 601 CTTCCAGTATAGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAACAATGGCTCATGCC 660  
DB 105647 CTTCCAGTATAGAAACCTTAAGATACCTAGAGCGGCTTTTGGGAACAATGGCTCATGCC 105706

QY 661 ACAGTAGTAGGACACATAATTGCTGCTGTATGGAATGTAATGGAATATGGATT 720  
DB 105707 ACAGTAGTAGGACACATAATTGCTGCTGTATGGAATGTAATGGAATATGGATT 105766  
  
QY 721 GC 722  
DB 105767 GC 105768  
  
RESULT 4  
US-09-949-016-16050  
; Sequence 16050, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 16050  
; LENGTH: 131379  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-16050

Query Match 98.9%; Score 714.8; DB 4; Length 131379;  
Best Local Similarity 99.2%; Pred. No. 1.8e-224;  
Matches 716; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 TGGAGCTGTCATGTTACCGTCTCTAAGCTTGGACTCTTAAGAAATGATTATTCCTGG 60  
DB 105047 TGGAGCTGTCATGTTACCGTCTCTAAGCTTGGACTCTTAAGAAATGATTATTCCTGG 105106  
  
QY 61 TTCTAGACAGCCAAATGTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 120  
DB 105107 TTCTAGACAGCCAAATGTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 105166  
  
QY 121 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 180  
DB 105167 ATTTGATTGGGTATTGAGCATGCTCTGAATGACAGTCCCCAAAAAGGACCTCTTATCCGT 105226  
  
QY 181 TCTTCCCTTGGGGAAGGCTTTTGGCACTTCCATGTCATATGTCAGTGGCCTTGAA 240  
DB 105227 TCTTCCCTTGGGGAAGGCTTTTGGCACTTCCATGTCATATGTCAGTGGCCTTGAA 105286  
  
QY 241 ATTTGATGCTGTACACATAGCAATTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 300  
DB 105287 ATTTGATGCTGTACACATAGCAATTAATTCACCTAGCTGACAGTAAAGAGTGGCTTACTAG 105346  
  
QY 301 ATAGATTCAAACCTAGTACTATGTGGACAGGGGGGCGACAGGACCCCACTTTGTAA 360  
DB 105347 ATAGATTCAAACCTAGTACTATGTGGACAGGGGGGCGACAGGACCCCACTTTGTAA 105406  
  
QY 361 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTTCTTATTTACCTGGCAAAATATCCAG 420  
DB 105407 AACATGTTTGGGGAATGTTTGTCTTATTTTCAATTTTCTTATTTACCTGGCAAAATATCCAG 105466  
  
QY 421 GTGGTGTGTAGTCCAGTAGAGATTATAAGTCCAGAGGAGTAGAATCAGCCTTACAA 480  
DB 105467 GTGGTGTGTAGTCCAGTAGAGATTATAAGTCCAGAGGAGTAGAATCAGCCTTACAA 105526  
  
QY 481 ACAGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCCAAGGAGCGAACTCAAAAGA 540  
DB 105527 ACAGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCCAAGGAGCGAACTCAAAAGA 105586





```

; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PT39pt-Fls
; US-08-232-463-14

Query Match 5.8%; Score 41.8; DB 1; Length 7218;
Best Local Similarity 1.6%; Pred No. 0.012;
Matches 5; Conservative 187; Mismatches 125; Indels 0; Gaps 0;

QY 406 GCGAAATATCCAGGTGGTGTGTGAGTCACAGCTAGAGATTATAAGTCCAAAGGAAGTA 465
Db 1378 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1319
QY 466 GAATCAGCCTTACAACAGTGGACCTCAACAAGAGAGATGTCGACCTGAACCCACWGAA 525
Db 1318 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1259
QY 526 GCGAAACTCAAAGAGAAAGAAAGCAACCAAGAACCTCTCGATGCGTTTCTCAGACAA 585
Db 1258 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1199
QY 586 ATGCTAAGCCCTTACTTCAGTATAGGAAACCTAAGATACCTAGACGGCTTTTCGGAA 645
Db 1198 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1139
QY 646 CAATGGCTCATGCCACAGGTAGTAGGACAGACATAATTGAGCTGGTGTATGGAATGTG 705
Db 1138 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1079
QY 706 AATGGAATATGGATTGC 722
Db 1078 RRRRRRRRRRATCGC 1062

RESULT 10
US-09-248-796A-3540
; Sequence 3540, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 3540
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (157), (170), (175)
; OTHER INFORMATION: Identity of nucleotide sequences at the above locations are unknown
; US-09-248-796A-3540

```

```
Query Match          5.0%; Score 36; DB 4; Length 1728;
Best Local Similarity 55.9%; Pred. No. 0.44;
Matches 66; Conservative 1; Mismatches 51; Indels 0; Gaps 0;

QY 439 GTAGAGATTATAAGTCCAGGAGTAGAATCAGCCTTCAACACAGTGACCTCAACGAA 498
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 GAAGTGATTGAATTTCTGATGATGTACAAAGCCACAGAACCAAGAAATCAACAA 930

QY 499 GGAAGTGTGCTCACTCAACCCGAGCGAAGCGAACTCAAAAGAGAAAGCAACCAA 556
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 GTAATTGAAGAAATTCGAAGAACTGAAGAACTGAAGAAATTCGAAGAAATTCGAAGAA 988

RESULT 11
US-09-949-016-162024/c
; Sequence 162024, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 162024
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-162024

Query Match          4.8%; Score 34.8; DB 4; Length 601;
Best Local Similarity 57.3%; Pred. No. 0.59;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 352 ACTTTGTAACAACTGTTTTGGGGAATGTTTTTTCATTTTCTTATTACCTGCGAAA 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 AGTGTGTGTACTGTGAATCGTCTATTTTATCAATGAAACATTTATAAAGTGAATA 389

QY 412 ATAATCCAGTGTGTGTGAGTCACAGTAGAGATTATAAGTCCAAAGGA 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 ATAATGCCCTTAATGTGAGGGTTTGTAAATGGTCTTATTAAAGACCAAGA 339

RESULT 12
US-09-949-016-204212
; Sequence 204212, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 204212
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-204212
```

```
; ORGANISM: Human
US-09-949-016-204212

Query Match          4.8%; Score 34.8; DB 4; Length 601;
Best Local Similarity 50.6%; Pred. No. 0.59;
Matches 81; Conservative 1; Mismatches 78; Indels 0; Gaps 0;

QY 523 GAAGCGAAACTCAAAAGAGAGAAAGCAACCAAGAACCTCTCTGATGCGTTTCTCAGA 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5 GGGCGAGACTCATCTCAAAAAAACAACAAAAAACAACAAACAGTGTGTGG 64

QY 583 CAAATGGTAAAGCCCTTACTTCCAGTATAGAAAACTTAAGATACCTTAGAGCGGCTTTGG 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 65 AGAAATCTGAACCCCTTGTGCGCTATTGGTGGGAATGTAATAATGGAGTAGCAGCTATGA 124

QY 643 GAACATGGGCTCATGCCACAGTAGTAGGACATAATT 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 125 AAACAGTGTGTCATTTCTCAAAATCCTGGACATAGAATT 164

RESULT 13
US-09-949-016-4532
; Sequence 4532, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4532
; LENGTH: 9668
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-4532

Query Match          4.8%; Score 34.8; DB 4; Length 9668;
Best Local Similarity 57.3%; Pred. No. 3;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 352 ACTTTGTAACAACTGTTTTGGGGAATGTTTTTTCATTTTCTTATTACCTGCGAAA 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9335 AGTGTGTGTACTGTGAATCGTCTATTTTATCAATGAAACATTTATAAAGTGAATA 9394

QY 412 ATAATCCAGTGTGTGTGAGTCACAGTAGAGATTATAAGTCCAAAGGA 461
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9395 ATAATGCCCTTAATGTGAGGGTTTGTAAATGGTCTTATTAAAGACCAAGA 9444

RESULT 14
US-09-949-016-17508/c
; Sequence 17508, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
```

Search completed: May 5, 2005, 07:03:13  
Job time : 170.759 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2005, 03:59:12 ; Search time 993.612 Seconds  
(without alignments)  
4449.254 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGCTTACC.....TGAATGGATATGATTGCG 723

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters: 11308400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	722.2	99.9	723	8	US-08-731-499-2
2	167.6	23.2	349	13	Sequence 899, App
3	167.6	23.2	2813	17	US-10-040-739-889
4	167.6	23.2	2813	17	Sequence 937, App
5	167.6	23.2	2813	17	US-10-172-118-937
6	167.6	23.2	2813	17	US-10-240-425-385
7	167.6	23.2	2813	17	US-10-342-887-937
8	167.6	23.2	2813	18	US-10-723-860-1812
9	167.6	23.2	3066	8	US-08-731-499-12
10	167.6	23.2	3418	18	US-10-723-860-6167
11	166	23.0	2020	19	US-10-491-213-85
12	88.6	12.3	383	9	US-09-960-352-10031
13	41.2	5.7	957	9	US-09-938-842A-845

12	41.2	5.7	957	11	US-09-938-842A-845	Sequence 845, App
13	39.4	5.4	103464	18	US-10-322-281-323	Sequence 323, App
14	39	5.4	3673778	16	US-10-312-841-1	Sequence 1, Appl
15	37.4	5.2	408	18	US-10-275-323A-13	Sequence 13, Appl
16	37.2	5.1	386	18	US-10-425-115-134418	Sequence 134418,
17	37.2	5.1	17534	17	US-10-257-166-108	Sequence 108, App
18	37.2	5.1	83836	13	US-10-087-192-1849	Sequence 1849, Ad
19	36.6	5.1	3353	18	US-10-357-930-25269	Sequence 25269, A
20	36.6	5.1	3353	18	US-10-357-930-27819	Sequence 27819, A
21	36.4	5.0	559	13	US-10-027-632-186364	Sequence 186364,
22	36.4	5.0	559	13	US-10-027-632-186365	Sequence 186365,
23	36.4	5.0	559	17	US-10-027-632-186364	Sequence 186364,
24	36.4	5.0	559	17	US-10-027-632-186365	Sequence 186365,
25	36.4	5.0	1113	13	US-10-027-632-260191	Sequence 260191,
26	36.4	5.0	1113	17	US-10-027-632-260191	Sequence 260191,
27	36.2	5.0	198522	13	US-10-087-192-244	Sequence 244, App
28	36	5.0	598	16	US-10-029-386-451	Sequence 451, App
29	36	5.0	885	17	US-10-371-264-83	Sequence 83, Appl
30	36	5.0	885	17	US-10-371-099-379	Sequence 379, App
31	36	5.0	885	17	US-10-371-122-379	Sequence 379, App
32	36	5.0	885	17	US-10-373-567-83	Sequence 83, Appl
33	36	5.0	885	17	US-10-628-088-379	Sequence 379, App
34	36	5.0	885	19	US-10-831-780-379	Sequence 379, App
35	36	5.0	13215	17	US-10-371-264-96	Sequence 96, Appl
36	36	5.0	13215	17	US-10-371-099-20	Sequence 20, Appl
37	36	5.0	13215	17	US-10-373-567-96	Sequence 96, Appl
38	36	5.0	13215	17	US-10-373-567-96	Sequence 96, Appl
39	36	5.0	13215	17	US-10-628-088-20	Sequence 20, Appl
40	36	5.0	13215	19	US-10-831-780-20	Sequence 20, Appl
41	36	5.0	13335	18	US-10-789-400-1	Sequence 1, Appli
42	36	5.0	14083	18	US-10-789-400-3	Sequence 3, Appli
43	35.8	5.0	56339	17	US-10-085-117-280	Sequence 280, App
44	35.4	4.9	811	14	US-10-184-644-414	Sequence 414, App
45	35.4	4.9	811	14	US-10-184-634-414	Sequence 414, App

#### ALIGNMENTS

RESULT 1  
US-08-731-499-2  
; Sequence 2, Application US/08731499  
; Publication No. US20030148270A1  
; GENERAL INFORMATION:  
; APPLICANT: GRAY, Joe W.  
; APPLICANT: COLLINS, Colin  
; APPLICANT: HWANG, Soo-In  
; APPLICANT: GODFREY, Tony  
; APPLICANT: KOWBEL, David  
; APPLICANT: ROMMENS, Johanna  
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR  
; TITLE OF INVENTION: USES  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/731,499  
; FILING DATE: 16-OCT-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/680,395  
; FILING DATE: 15-JUL-1996  
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 23070-068910
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..723
; OTHER INFORMATION: /note="cDNA clone lb11 of 3.5kb
; OTHER INFORMATION: transcript"
US-08-731-499-2

Query Match
Best Local Similarity 100.0%; Score 722.2; DB 8; Length 723;
Matches 723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAAGCTGTCATGTTACCGTCTCTAAAGCTTGGACTCTTAAGAAATGATTATTCCTGG 60
DB |||||
QY 1 TGAAGCTGTCATGTTACCGTCTCTAAAGCTTGGACTCTTAAGAAATGATTATTCCTGG 60
DB |||||
QY 61 TTCTTAGACAGGCCAAATGTAATCACTACGTGGCAGATTAAGAGGTGGGCTTACTAG 120
DB |||||
QY 61 TTCTTAGACAGGCCAAATGTAATCACTACGTGGCAGATTAAGAGGTGGGCTTACTAG 120
DB |||||
QY 121 ATTGATTTGGTATTGAGCATGCTCTGAATGACAGTCCCAAGAGGACCTCTTATCCGT 180
DB |||||
QY 121 ATTGATTTGGTATTGAGCATGCTCTGAATGACAGTCCCAAGAGGACCTCTTATCCGT 180
DB |||||
QY 181 TCTTCCCTTGGGGAAGGGCTTTGGCAGTCTTCAATGTCAGTGGCAGTTGAGCTTGGAA 240
DB |||||
QY 181 TCTTCCCTTGGGGAAGGGCTTTGGCAGTCTTCAATGTCAGTGGCAGTTGAGCTTGGAA 240
DB |||||
QY 241 ATTGTCGCTGTGACAACTAAGATTAATCACTTCAAGATGTCCTGTGTAGAAATGGTC 300
DB |||||
QY 241 ATTGTCGCTGTGACAACTAAGATTAATCACTTCAAGATGTCCTGTGTAGAAATGGTC 300
DB |||||
QY 301 ATAGATTCAAACTGTAGTACTATGTGGACGGGGGCGCAGCAGACCCCACTTTGTA 360
DB |||||
QY 301 ATAGATTCAAACTGTAGTACTATGTGGACGGGGGCGCAGCAGACCCCACTTTGTA 360
DB |||||
QY 361 AACATGTTTGGGGGAATGTTTGTGTTTCTTATTTCTTATTTCTGCAAAATAATCCAG 420
DB |||||
QY 361 AACATGTTTGGGGGAATGTTTGTGTTTCTTATTTCTTATTTCTGCAAAATAATCCAG 420
DB |||||
QY 421 GTGCTGTGTGAGTCACAGTAGAGATTATAAGTCCAAGGAAGTAGAATCAGCCTTACAA 480
DB |||||
QY 421 GTGCTGTGTGAGTCACAGTAGAGATTATAAGTCCAAGGAAGTAGAATCAGCCTTACAA 480
DB |||||
QY 481 ACAGTGGACCTCAACGAGGAGATGTCGACCTGAACCCACGACGGAAGGAACTCAAAGA 540
DB |||||
QY 481 ACAGTGGACCTCAACGAGGAGATGTCGACCTGAACCCACGACGGAAGGAACTCAAAGA 540
DB |||||
QY 541 GAAGAAGCAAAACCAAGAACTCTCTGATGCGCTTTCTCAGACAAATGTTAAGCCCTTA 600
DB |||||
QY 541 GAAGAAGCAAAACCAAGAACTCTCTGATGCGCTTTCTCAGACAAATGTTAAGCCCTTA 600
DB |||||
QY 601 CTTTCCAGTATAGGAAACCTAAGATACCTAGACGGGCTTTTGGAAACAAATGGGCTCATGCC 660
DB |||||
QY 601 CTTTCCAGTATAGGAAACCTAAGATACCTAGACGGGCTTTTGGAAACAAATGGGCTCATGCC 660
DB |||||
QY 661 ACAGGTAGTAGGACACATAATTGTAGCTGTGTATGGAATCTGGAATATGGATT 720
DB |||||
QY 661 ACAGGTAGTAGGACACATAATTGTAGCTGTGTATGGAATCTGGAATATGGATT 720
DB |||||
QY 721 GCG 723

```

```

DB 721 GCG 723
|||

RESULT 2
US-10-040-739-889
; Sequence 889, Application US/10040739
; Publication No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; Lavallie, Edward
; Racie, Lisa
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
; NUMBER OF SEQUENCES: 1519
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/040,739
; FILING DATE: 07-Jan-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/036,520
; FILING DATE: 03-JUN-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Scott A.
; REGISTRATION NUMBER: 32,724
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 889:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 889:
US-10-040-739-889

Query Match
Best Local Similarity 98.2%; Pred. No. 1.2e-40;
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCACAGTAGAGATTATAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
DB |||||
QY 419 AGTGTGTGTGAGTCACAGTAGAGATTATAAGTCCAAGGAAGTAGAATCAGCCTTAC 478
DB |||||
QY 479 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCACGAAAGGAACTCAAAA 538
DB |||||
QY 129 AAACAGTGGACCTCAACGAGGAGATGCTGCACCTGAACCCACGAAAGGAACTCAAAA 188
DB |||||
QY 539 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 588
DB |||||
QY 189 GAGAAGAAAGCAAAACCAAGAACCTCTCTGATGCGTTTCTCAGACAAATG 238
DB |||||

RESULT 3
US-10-172-118-937
; Sequence 937, Application US/10172118

```

; Publication No. US20030224374A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dai, Hongyue  
 ; APPLICANT: He, Yudong  
 ; APPLICANT: Linsley, Peter  
 ; APPLICANT: Mao, Mao  
 ; APPLICANT: Roberts, Chris  
 ; APPLICANT: Van 't Veer, Laura  
 ; APPLICANT: Van de Vijver, Marc  
 ; APPLICANT: Bernards, Rene  
 ; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients  
 ; FILE REFERENCE: 9301-175-999  
 ; CURRENT APPLICATION NUMBER: US/10/172,118  
 ; CURRENT FILING DATE: 2002-06-14  
 ; PRIOR APPLICATION NUMBER: 60/380,770  
 ; PRIOR FILING DATE: 2002-05-14  
 ; NUMBER OF SEQ ID NOS: 2699  
 ; SEQ ID NO 937  
 ; LENGTH: 2813  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; PUBLICATION INFORMATION:  
 ; DATABASE ACCESSION NUMBER: NM\_003657  
 ; DATABASE ENTRY DATE: 2001-06-18  
 ; US-10-172-118-937  
  
 Query Match 23.2%; Score 167.6; DB 17; Length 2813;  
 Best Local Similarity 99.2%; Pred. No. 3.4e-40;  
 Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
 QY 419 AGGTGGGTGTGAGTCCACGAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538  
 DB 1298 ATGTGGGTGTGAGTCCACGAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1417  
  
 QY 479 AAACAGTGGACCTCAACGAAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 538  
 DB 1358 AAACAGTGGACCTCAACGAAAGGAGATGCTGCACCTGAAACCCACGAAAGCGAAACTCAAAA 1417  
  
 QY 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588  
 DB 1418 GAGAAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 1467  
  
 RESULT 4  
 US-10-240-425-385  
 ; Sequence 385, Application US/10240425  
 ; Publication No. US20040033502A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Williams, Amanda  
 ; APPLICANT: Boland, Joseph F.  
 ; APPLICANT: Lord, Reginald V.  
 ; APPLICANT: Alvarez, Chris  
 ; APPLICANT: Wetzel, Jon C.  
 ; APPLICANT: Scherf, Uwe  
 ; APPLICANT: Vockley, Joseph G.  
 ; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue  
 ; FILE REFERENCE: 44921-5026  
 ; CURRENT APPLICATION NUMBER: US/10/240,425  
 ; PRIOR FILING DATE: 2002-09-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/09847  
 ; PRIOR FILING DATE: 2001-03-28  
 ; PRIOR APPLICATION NUMBER: US 60/193,446  
 ; PRIOR FILING DATE: 2000-03-31  
 ; NUMBER OF SEQ ID NOS: 1588  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 385  
 ; LENGTH: 2813  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AF041260  
 ; US-10-240-425-385

/ TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
/ FILE REFERENCE: 05882.0193.NPUS01  
/ CURRENT APPLICATION NUMBER: US/10/723,860  
/ CURRENT FILING DATE: 2003-11-26  
/ PRIOR APPLICATION NUMBER: 60/429,739  
/ PRIOR FILING DATE: 2002-11-26  
/ NUMBER OF SEQ ID NOS: 8393  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 1812  
/ LENGTH: 2813  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
US-10-723-860-1812

Query Match 23.2%; Score 167.6; DB 18; Length 2813;  
Best Local Similarity 98.2%; Pred. No. 3.4e-40;  
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTGAGTCACCAAGGAGATTTATAAGTCCAAGGAGTAGAATCAGCCTTAC 478  
Db 1298 ATGTGGTGTGTGAGTCACCAAGGAGATTTATAAGTCCAAGGAGTAGAATCAGCCTTAC 1357

QY 479 AACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAGGAGAACTCAAAA 538  
Db 1358 AACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAGGAGAACTCAAAA 1417

QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 588  
Db 1418 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 1467

RESULT 7  
US-08-731-499-12  
/ Sequence 12, Application US/08/731499  
/ Publication No. US20030148270A1  
/ GENERAL INFORMATION:  
/ APPLICANT: GRAY, Joe W.  
/ APPLICANT: COLLINS, Colin  
/ APPLICANT: HWANG, Soo-In  
/ APPLICANT: GODFREY, Tony  
/ APPLICANT: KOMBEL, David  
/ APPLICANT: ROMMENS, Johanna  
/ TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR  
/ NUMBER OF SEQUENCES: 44  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Townsend and Townsend and Crew  
/ STREET: Two Embarcadero Center, 8th Floor  
/ CITY: San Francisco  
/ STATE: California  
/ COUNTRY: USA  
/ ZIP: 94111-3834  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/731,499  
/ FILING DATE: 16-OCT-1996  
/ CLASSIFICATION: 435  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 08/680,395  
/ FILING DATE: 15-JUL-1996  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Hunter, Tom  
/ REGISTRATION NUMBER: 38,498  
/ REFERENCE/DOCKET NUMBER: 23070-068910  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (415) 576-0200  
/ TELEFAX: (415) 576-0300  
/ INFORMATION FOR SEQ ID NO: 12:  
/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 3066 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: cDNA  
/ FEATURE:  
/ NAME/KEY: -  
/ LOCATION: 1..3066  
/ OTHER INFORMATION: /note= "1b1"  
US-08-731-499-12

Query Match 23.2%; Score 167.6; DB 8; Length 3066;  
Best Local Similarity 98.2%; Pred. No. 3.6e-40;  
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTGAGTCACCAAGGAGATTTATAAGTCCAAGGAGTAGAATCAGCCTTAC 478  
Db 1433 ATGTGGTGTGTGAGTCACCAAGGAGATTTATAAGTCCAAGGAGTAGAATCAGCCTTAC 1492

QY 479 AACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAGGAGAACTCAAAA 538  
Db 1493 AACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAGGAGAACTCAAAA 1552

QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 588  
Db 1553 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 1602

RESULT 8  
US-10-723-860-6167  
/ Sequence 6167, Application US/10723860  
/ Publication No. US20040253606A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Aziz, Natasha  
/ APPLICANT: Ginsburg, Wendy M.  
/ APPLICANT: Zlotnik, Albert  
/ TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
/ FILE REFERENCE: 05882.0193.NPUS01  
/ CURRENT APPLICATION NUMBER: US/10/723,860  
/ CURRENT FILING DATE: 2003-11-26  
/ PRIOR APPLICATION NUMBER: 60/429,739  
/ PRIOR FILING DATE: 2002-11-26  
/ NUMBER OF SEQ ID NOS: 8393  
/ SOFTWARE: PatentIn version 3.2  
/ SEQ ID NO 6167  
/ LENGTH: 3418  
/ TYPE: DNA  
/ ORGANISM: Homo sapiens  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (1815)..(1815)  
/ OTHER INFORMATION: n is a, c, g, or t  
/ FEATURE:  
/ NAME/KEY: misc feature  
/ LOCATION: (3303)..(3416)  
/ OTHER INFORMATION: n is a, c, g, or t  
US-10-723-860-6167

Query Match 23.2%; Score 167.6; DB 18; Length 3418;  
Best Local Similarity 98.2%; Pred. No. 3.8e-40;  
Matches 167; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTGAGTCACCAAGGAGATTTATAAGTCCAAGGAGTAGAATCAGCCTTAC 478  
Db 1346 ATGTGGTGTGTGAGTCACCAAGGAGATTTATAAGTCCAAGGAGTAGAATCAGCCTTAC 1405

QY 479 AACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAGGAGAACTCAAAA 538  
Db 1406 AACAGTGGACCTCAACGAAGGAGATGCTGCACCTGAACCCACGAGGAGAACTCAAAA 1465

QY 539 GAGAGAAAGCAAAACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATG 588  
|||||



DB 1466 GAGAAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 1515

RESULT 9

US-10-491-213-85  
; Sequence 85, Application US/10491213  
; Publication No. US20050048490A1  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE CORPORATION; AZIMZAI, Yalda;  
; APPLICANT: BAUGHN, Mariah R.; BECHA, Shanya D.;  
; APPLICANT: BOWEN, Mark L.; CHAWLA, Narinder K.;  
; APPLICANT: ELLIOTT, Vicki S.; EMERLING, Brooke M.;  
; APPLICANT: GANDHI, Aneena R.; GIETZEN, Kimberly J.;  
; APPLICANT: GORVAD, Ann E.; GRIFFIN, Jennifer A.;  
; APPLICANT: HAFALIA, April J.A.; ISON, Craig H.;  
; APPLICANT: KABLE, Amy E.; KALAFUS, Daniel P.;  
; APPLICANT: LEHR-MASON, Patricia M.; LU, Dying Aina M.;  
; APPLICANT: MARQUIS, Joseph P.; NGUYEN, Dannel B.;  
; APPLICANT: RAMKUMAR, Jayalaxmi; RICHARDSON, Thomas W.;  
; APPLICANT: KAREHT, Stephanie K.; SWARNAKAR, Anita;  
; APPLICANT: TANG, Y. Tom; TRAN, Uyen K.;  
; APPLICANT: WARREN, Bridget A.; XU, Yuming;  
; APPLICANT: YAO, Monique G.; YUE, Huibin;  
; APPLICANT: YUE, Henry  
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL GROWTH, DIFFERENTIATION, AND DEATH  
; FILE REFERENCE: PF-1213 USN  
; CURRENT APPLICATION NUMBER: US/10/491,213  
; PRIOR FILING DATE: 2004-03-26  
; PRIOR APPLICATION NUMBER: PCT/US02/31095  
; PRIOR FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/326,389  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: US 60/327,380  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/328,186  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: US 60/329,690  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/345,384  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 60/348,165  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 60/350,219  
; PRIOR FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: US 60/344,518  
; PRIOR FILING DATE: 2001-11-09  
; PRIOR APPLICATION NUMBER: US 60/345,143  
; PRIOR FILING DATE: 2001-11-09  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 114  
; SOFTWARE: PERL Program  
; SEQ ID NO 85  
; LENGTH: 2020  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No: 4757551CB1

Query Match 23.0%; Score 166; DB 19; Length 2020;  
Best Local Similarity 97.6%; Pred. No. 9e-40;  
Matches 166; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 419 AGTGGTGTGTCAGTCACCGTAGAGATATTAAGTCCAGGAGTAGAATCAGCCTTAC 478  
DB 1312 ATGTGGTGTGTCAGTCACCGTAGAGATATTAAGTCCAGGAGTAGAATCAGCCTTAC 1371  
QY 479 AAACAGTGCACCTCAACGAAGAGAGTGTGACCTTGAACCCACGAGGAGAACTCAAAA 538  
DB 1372 AAACAGTGCACCTCAACGAAGAGAGTGTGCTCTGACCCACGAGGAGAACTCAAAA 1431  
QY 539 GAGAAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 588

DB 1432 GAGAAGAAAGCAACCAAGAACCTCTCTGATGGCGTTTCTCAGACAAATG 1481

RESULT 10

US-09-960-352-10031  
; Sequence 10031, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 10031  
; LENGTH: 383  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (336)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: 43-LIB34-002-Q1-E1-C12  
US-09-960-352-10031

Query Match 12.3%; Score 88.6; DB 9; Length 383;  
Best Local Similarity 62.9%; Pred. No. 2.3e-16;  
Matches 168; Conservative 2; Mismatches 91; Indels 6; Gaps 2;  
QY 328 GGACAGGGGGGCGACCAAGACCCACCTTTGTAACATGTTTGGGGAATGTTTGT 387  
DB 51 GAAAAGGG 110  
QY 388 TTCAATTTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCCACGATGAGATT 447  
DB 111 GAGGATTTCAGTCCCGACAGGTGCAGAGGAGAGATGTTGTGTGTGAGTCCACGATGAGGCT 170  
QY 448 ATAAAGTCCAGGAGTAGATCAGCTTACAAACAGTGGACCT---CAACGAGGAGAT 504  
DB 171 GTAAAGTCTGAAGAGTAGAGTACGCTTACAAACAGTGGATCTCGGCGAAGAGGAGAG 230  
QY 505 GCTGCCTCTGAACCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561  
DB 231 CGCACCCGATCCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290  
QY 562 TCTCTGATGTCGTTTCTCAGACAAATG 588  
DB 291 TCCCTGATGGCGCTCCTCAGACAGATG 317

RESULT 11

US-09-938-842A-845  
; Sequence 845, Application US/09938842A  
; Patent No. US20020160378A1  
; GENERAL INFORMATION:  
; APPLICANT: Harper, Jeff  
; APPLICANT: Kreps, Joel  
; APPLICANT: Wang, Xun  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING  
; FILE REFERENCE: SAME, AND METHODS OF USE  
; CURRENT APPLICATION NUMBER: US/09/938,842A  
; CURRENT FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: US 60/227,866  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: US 60/264,647  
; PRIOR FILING DATE: 2001-01-16

```

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 845
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-845

Query Match
Best Local Similarity 5.7%; Score 41.2; DB 9; Length 957;
Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;

QY 382 TTGTTTTTCATTTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCACCAGTA 441
Db 358 TGGATGTTTAGAGTCTTTGGACATGACAAAGTATGGTACTTGTATGGAGGCTTGCCGAA 417
QY 442 GAGATTATAAGTCCAGGAAGTAGAATCAGCCCTTACAAACAGTGAACCTCAACGAAGGA 501
Db 418 TGGCGTGTCTTCAGGCTATGACGTTGAATCCAGTGTTCAAATGATGCCATTTTGAAGGC 477
QY 502 GATGCTGCACCTGAACCCACGAGGAACTCAAAAGAGAAAGCAACCAACGAACC 561
Db 478 AGTGCAGCAACTGAGGCTATAGAAAGATCTATCAAGGACAAACAATTAGCCCAATAACT 537
QY 562 TCTCTGATGRCGTTTCTCAGACAAATGGTAAGCCCTTACTTCCAGTATAGGAACCTAA 621
Db 538 TTTCAAGCAAGTTCGGCCACATCTTGTATTTGGCACTTGTATCAGGTGAAGGAACATC 597
QY 622 GATACCTAGA 631
Db 598 GAGGATAAGA 607

RESULT 12
US-09-938-842A-845
; Sequence 845, Application US/09938842A
; Publication No. US20040009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 845
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-845

Query Match
Best Local Similarity 5.7%; Score 41.2; DB 11; Length 957;
Matches 118; Conservative 2; Mismatches 130; Indels 0; Gaps 0;

QY 382 TTGTTTTTCATTTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTGTGAGTCACCAGTA 441
Db 358 TGGATGTTTAGAGTCTTTGGACATGACAAAGTATGGTACTTGTATGGAGGCTTGCCGAA 417
QY 442 GAGATTATAAGTCCAGGAAGTAGAATCAGCCCTTACAAACAGTGAACCTCAACGAAGGA 501
Db 418 TGGCGTGTCTTCAGGCTATGACGTTGAATCCAGTGTTCAAATGATGCCATTTTGAAGGC 477

; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 845
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-845

Query Match
Best Local Similarity 5.7%; Score 39.4; DB 18; Length 103464;
Matches 85; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 107 GGTGGGCTTACTAGATTGATTTGGTGGTATTGAGCATGCTCTGAATGACACATCCCCAAAAG 166
Db 94634 GATGGGCTTTCATATGTATATTTGGCAAAATTCATGTAATAGAAACAATTTAGAAAGAG 94575
QY 167 GACCTCTTATCCGTTCTTCCCTCTGGGGAAGGCTTTTCCCACTTCCATGTCAATGTGGC 226
Db 94574 AAGGTTTATAGGTTTATTTATTGAGAAATGTTTTTATAGATACAAGTAATGTAGT 94515
QY 227 AGTTGAGCTTGGAAATTTGGTGGCTTTGTACAACATAAGCATT 267
Db 94514 TTTAAAAATTATGACTTTTGAAGGTAGTAGGACATAAAAAAT 94474

RESULT 14
US-10-312-841-1/c
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: EpiGenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-322-281-323/c
; Sequence 323, Application US/10322281
; Publication No. US20040126762A1
; GENERAL INFORMATION:
; APPLICANT: David W. Morris
; APPLICANT: Marc S. Malandro
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer
; FILE REFERENCE: 529452001000
; CURRENT APPLICATION NUMBER: US/10/322,281
; CURRENT FILING DATE: 2002-12-17
; NUMBER OF SEQ ID NOS: 866
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 323
; LENGTH: 103464
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(103464)
; OTHER INFORMATION: n = A,T,C or G
US-10-322-281-323
```

US-10-312-841-1

Query Match	5.4%	Score 39	DB 16	Length 3673778
Best Local Similarity	48.8%	Pred. No. 44		
Matches 102	Conservative 1	Mismatches 106	Indels 0	Gaps 0
Qy	410	AAATAATCCAGTGGTGTGTGAGTCACAGTAGAGATTATAAAGTCCCAAGGAAGTAGAAT	469	
Db	1519348	AAATACTAATATTATTTTTTTTAACAATCTTACAATATAATTACACCAAAATATAA	1519289	
Qy	470	CAGCCTTACAAACAGTGGACCTCAACGAAAGAGATGCTGCACCTGAACCCCAACWGAAGCGA	529	
Db	1519288	ATACCTAAATTTATATTTACATTTTTTAAACAAATAAATTTTTAAAAAATAATCCATATAAAAAA	1519229	
Qy	530	AACTCAAAAGAGAGAAAGCAACCAAGAACCTCTCTGATGRCGTTTTCTCAGACAAATGG	589	
Db	1519228	AAAACAAAATCAAAAAAANAACACACAAAATTTCCATACCAATCCTTAATACAAATAA	1519169	
Qy	590	TAAGCCCCCTTACTTCCAGTATAGAAACC	618	
Db	1519168	CAATCCCTTTACTATCATAAAAATTCACC	1519140	

RESULT 15  
US-10-275-323A-13/c  
Sequence 13, Application US/10275323A  
Publication No. US20050079491A1  
GENERAL INFORMATION:  
APPLICANT: DONNE-GOUSSE, CAROLE  
APPLICANT: LAUDET, VINCENT  
APPLICANT: HANNI, CATHERINE  
TITLE OF INVENTION: METHOD OF DETECTING AND IDENTIFYING THE PRESENCE OF BIOLOGICAL  
TITLE OF INVENTION: MATERIALS DERIVED FROM BIRDS, AND OLIGONUCLEOTIDES FOR CARRYING  
TITLE OF INVENTION: IT OUT  
FILE REFERENCE: 0508-1052  
CURRENT APPLICATION NUMBER: US/10/275.323A  
CURRENT FILING DATE: 2002-11-08  
PRIOR APPLICATION NUMBER: PCT/FR01/01279  
PRIOR FILING DATE: 2001-04-26  
PRIOR APPLICATION NUMBER: FR 00/05850  
PRIOR FILING DATE: 2000-05-09  
NUMBER OF SEQ ID NOS: 96  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 13  
LENGTH: 408  
TYPE: DNA  
ORGANISM: Artificial  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: DNA fragment amplified  
OTHER INFORMATION: from the genome of birds  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (43)..(43)  
OTHER INFORMATION: a, c, g, or t  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (57)..(57)  
OTHER INFORMATION: a, c, g, or t  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (71)..(71)  
OTHER INFORMATION: a, c, g, or t  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (134)..(134)  
OTHER INFORMATION: a, c, g, or t  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (212)..(212)  
OTHER INFORMATION: a, c, g, or t  
FEATURE:  
NAME/KEY: modified base  
LOCATION: (214)..(214)

1	OTHER INFORMATION: a, c, g, or t
2	FEATURE:
3	NAME/KEY: modified base
4	LOCATION: (217)..(217)
5	OTHER INFORMATION: a, c, g, or t
6	FEATURE:
7	NAME/KEY: modified base
8	LOCATION: (226)..(227)
9	OTHER INFORMATION: a, c, g, or t
10	FEATURE:
11	NAME/KEY: modified base
12	LOCATION: (236)..(236)
13	OTHER INFORMATION: a, c, g, or t
14	FEATURE:
15	NAME/KEY: modified base
16	LOCATION: (245)..(245)
17	OTHER INFORMATION: a, c, g, or t
18	FEATURE:
19	NAME/KEY: modified base
20	LOCATION: (253)..(253)
21	OTHER INFORMATION: a, c, g, or t
22	FEATURE:
23	NAME/KEY: modified base
24	LOCATION: (269)..(269)
25	OTHER INFORMATION: a, c, g, or t
26	FEATURE:
27	NAME/KEY: modified base
28	LOCATION: (324)..(324)
29	OTHER INFORMATION: a, c, g, or t
30	FEATURE:
31	NAME/KEY: modified base
32	LOCATION: (330)..(330)
33	OTHER INFORMATION: a, c, g, or t
34	FEATURE:
35	NAME/KEY: modified base
36	LOCATION: (352)..(352)
37	OTHER INFORMATION: a, c, g, or t
38	FEATURE:
39	NAME/KEY: modified base
40	LOCATION: (355)..(355)
41	OTHER INFORMATION: a, c, g, or t
42	FEATURE:
43	NAME/KEY: modified base
44	LOCATION: (357)..(357)
45	OTHER INFORMATION: a, c, g, or t
46	FEATURE:
47	NAME/KEY: modified base
48	LOCATION: (362)..(362)
49	OTHER INFORMATION: a, c, g, or t
50	FEATURE:
51	NAME/KEY: modified base
52	LOCATION: (364)..(364)
53	OTHER INFORMATION: a, c, g, or t
54	FEATURE:
55	NAME/KEY: modified base
56	LOCATION: (406)..(406)
57	OTHER INFORMATION: a, c, g, or t
58	US-10-275-323A-13

Query Match	5.2%	Score 37.4;	DB 19;	Length 408;
Best Local Similarity	9.4%;	Pred. No. 1.2;		
Matches	36;	Conservative 204;	Mismatches 140;	Indels 1; Gaps 1;

  

Qy	92	GTGGCAGATTAAAGAGTGGGCTTACTAGATTTGATTTGGGTATTGACGATGCTCTGAATG	151
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
Db	399	SWAKGHRRYRDDBRDKYVRYTRRYDRDHYRDRBNVDDKTNKRANRBDRRDRTWRT	340
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
Qy	152	ACAGTCCCCAAAAGGACCTCTTATCGTTCCTCCCTTGCGGAAGGGCTTTTGCACACTT	211
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
Db	339	RKDWDYYNRDKTDNDWBHDBBYYTRGTCTCGWGTTYCTMTAATAADKWTRDDHRHD	280
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
Qy	212	CCAATGCAATGTGGCAGTTGAGCTTGGAANAATGGTGGCTTGTACAACATTAAGCATTA	271
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:
	:	: :   : : : : : :	:

279	Db	HHRTKDHWYNNKKHGDHWHKHKHVYVNDDBVWNRDDWRSRNRWTDWDDDRNDGVERT	220
272	Qy	CTCAAGATGTGCCTG-TGTAGAAATGGTCATAGATTCAAACTAGTACTACTATGTGGA	330
219	Db	WNRDNDNDDDWNTVDVKWDKDDDDKDBDRDWWKAYRHWABRWDKDDWHDWDAAW	160
331	Qy	CAGGGGGCAGCAGGACCCACATTGTGTAACATGTTTTGGGGAAATGTTTGTGTTTC	390
159	Db	YDGHKRAARWHYDTGTGTTKHTTNTKDDDKYSYDRRRRWYRSWDRWRDKWDWTCK	100
391	Qy	ATTTTCTTATTACCTGGCAAAATAATCCAGGTGTGTGTAGTCCACGATGAGATTATA	450
99	Db	WTWRHRTAATGTCDDCFAGCATTCACYNRYARCAYBWDVNVTRGDBDGYGGAQRK	40
451	Qy	AAGTCCAAGGAAGTAGAATCA	471
39	Db	RCMWYAAACYRAATGHRHCMA	19

Search completed: May 5, 2005, 12:12:19  
Job time : 1007.61 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:17:50 ; Search time 3060.29 Seconds  
(without alignments)  
8992.766 Million cell updates/sec

Title: US-08-731-499-2

Perfect score: 723

Sequence: 1 TGGAGCTGTCATGCTTACC.....TGAATGGAATGATTGCG 723

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_gesi:\*

9: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	384.2	53.1	504	5	BQ320066 PM3-CT081
C 2	169.2	23.4	5738	3	CR749643 Homo sapi
C 3	166	23.0	280	4	BI041215 IL5-NT027
C 4	166	23.0	837	4	BG290366 602388204
C 5	127.6	17.6	545	8	BH086142 RPCI-24-3
C 6	123	17.0	381	2	AW453053 UI-H-BW1
C 7	122.2	16.9	394	1	AA644590 af74a01.r
C 8	104.2	14.4	416	1	AI060049 UI-R-C1-k
C 9	100	13.8	358	1	AI459809 ap17h05.x
C 10	98.8	13.7	583	6	CB586288 AMGNNUC.N
C 11	89.8	12.4	359	2	BE485778
C 12	88.2	12.2	246	2	BE477506
C 13	86.6	12.0	668	4	BG964319 602831931
C 14	86.6	12.0	892	5	BU841356 AGENCOURT
C 15	86.6	12.0	2339	3	AK008957 Mus muscu
C 16	81.4	11.3	906	2	BE305008
C 17	66.2	9.2	902	2	BF581751 602099333
C 18	51.4	7.1	1000	2	BF577458 602092302
C 19	45.6	6.3	322	2	BF853823 MR2-EN009
C 20	43.8	6.1	681	6	CB423181 596383 MA
C 21	42.8	5.9	1341	3	CB817048 Arabidops
C 22	42.2	5.8	1208	5	BX988592
C 23	41.6	5.8	879	2	BE740812
C 24	41.4	5.7	922	4	BG961911 602826590

C 25	41.2	5.7	591	8	AQ934160	AQ934160	RPCI-23-2
C 26	41.2	5.7	755	5	B0635914	B0635914	042F07 In
C 27	41.2	5.7	1256	3	CNSOACXK	BX814914	Arabidops
C 28	40.6	5.6	1111	5	BX442513	BX442513	
C 29	40	5.5	391	7	CO954055	CO954055	UNC-pnata
C 30	39.8	5.5	282	1	AJ671212	AJ671212	
C 31	39.8	5.5	1057	9	CL108034	CL108034	ISB1-49D4
C 32	39.6	5.5	1163	2	BF691065	BF691065	60247118
C 33	39.4	5.4	635	4	BI390256	BI390256	pppic.pk0
C 34	39.4	5.4	734	9	CE248517	CE248517	tigr-gss-
C 35	39.2	5.4	600	7	CO879260	CO879260	BovGen.07
C 36	39.2	5.4	684	9	BI133486	BI133486	Danio rer
C 37	39	5.4	615	7	CN034417	CN034417	Math.pl.D
C 38	38.6	5.3	535	8	BH402050	BH402050	AG-ND-139
C 39	38.6	5.3	722	7	CN216393	CN216393	30290 Sus
C 40	38.6	5.3	755	8	BH723558	BH723558	BOMJV18TF
C 41	38.6	5.3	1302	5	BU505079	BU505079	AGENCOURT
C 42	38.4	5.3	658	9	BI172820	BI172820	Danio rer
C 43	38.4	5.3	722	7	CN059701	CN059701	Salamande
C 44	38.2	5.3	202	1	AA109272	AA109272	MBACE280
C 45	38.2	5.3	689	9	CE115637	CE115637	tigr-gss-

## ALIGNMENTS

LOCUS BQ320066 504 bp mRNA linear EST 17-MAY-2002  
DEFINITION PM3-CT0817-190501-005-d12 CT0817 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BQ320066  
VERSION BQ320066.1 GI:20926791  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 504)  
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,  
O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202653  
10737800  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM3&tr2=PM3-CT0817-  
190501-005-d12&t3=2001-05-19&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 433.  
Location/Qualifiers  
1..504  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="CT0817"  
/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived

# SUMMARIES



Location/Qualifiers  
1. .837  
/organism="Homo sapiens"

```
/clone_lib="RPCI-24"
/notes="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."
```

ORIGIN

Query Match	17.6%;	Score 127.6;	DB 8;	Length 645;	
Best Local Similarity	70.5%;	Pred. No. 2.8e-26;			
Matches 182;	Conservative	2;	Mismatches 71;	Indels 3;	Gaps 1;

QY 371 GGGGGAATGTTTGTGTTTTCATTTCTTATTACCTGGCAAAATAATCCAGGTGGTGTG 430  
|||||  
Db 318 GGGTGTGTTTGTGTTTTCATTTCTTATTACCTGGCAAAATAATCTAGCGGTGTG 377  
|||||

QY 431 AGTCACCATAGAGATTAATAGTCCAAGAGTAGAATCAGCTTACAAAAGTGGACC 490  
|||||  
Db 378 AGTCACCATAGAGACCGTAAGCTTGGAGAAAGTAGAATCCAGCTTACAAAAGTGGATC 437  
|||||

QY 491 TCAAGCAAGGAGATGCTGCACCTGACCCACACGACGAACTCAAAAGAGAGAAAGCA 550  
|||||  
Db 438 TCAGTGA---AGAGACCGACCTGAAACCCACACAGACGTAAGTCAAAAGAGAAAGCAAC 494  
|||||

QY 551 AACCAAGAACCTCTCTGATGTCGTTTCTCAGACAAATGGTAAGCCCTTACTTCCAGTAT 610  
|||||  
Db 495 CCGGGAAGACCTCTGATGGCTTCTCAGACAAATGGTAAGCCACTTCTGATGTTC 554  
|||||

QY 611 AGAAACCTAAGTACTT 628  
|||||  
Db 555 TGCAGAGGGGACATGCCT 572  
|||||

RESULT 6  
AW453053/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AW453053  
UI-H-BW1-ama-e-10-0-UI.61 NCI CGAP\_Sub7 Homo sapiens cDNA clone  
IMAGE:3069402 3', mRNA sequence.

AW453053.1 GI:6993829  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 381)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
Source

1..381  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3069402"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="NCI CGAP\_Sub7"  
/notes="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; NCI CGAP\_Sub7  
is a subtracted library derived from NCI CGAP\_Sub6. The  
NCI CGAP\_Sub7 library had 12 million recombinants. A

single-stranded DNA preparation of NCI CGAP\_Sub6 was used  
as a tracer in a subtractive hybridization with a driver  
comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM  
3334-3337, 3682-3683, 3798-3803 (IMAGE Clones)  
132376-132391, 145608-145677, 150552-150855);  
NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778  
(IMAGE Clones 1323912-1325831, 1471368-1472903,  
1492104-1493255); NCI CGAP Lu5 pool 1 LLAM 3575-3582,  
3851-3854 (IMAGE Clones 1414920-1417991,  
1520904-1522439); NCI CGAP\_G4 pool 1 LLAM 3164-3167,  
3716-3720, 3733-3735 (IMAGE Clones  
1257096-1258631, 1469064-1470983, 1475592-1476743);  
NCI CGAP\_P22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068  
(IMAGE Clones 985608-986759, 1101192-1101959,  
1217928-1220615); NCI CGAP\_Co10 pool 1 LLAM 2644-2653,  
2871-2872 (IMAGE Clones  
1057416-1061255, 1144584-1145351). (6% of the driver  
population), plus a pool of 3,840 arrayed clones from  
NCI CGAP\_Sub1 (IMAGE Clones 2708616-2710535) and  
NCI CGAP\_Sub2 (IMAGE Clones 2710536-2712455) (4% of  
the driver population), plus a pool of 11,136 clones from  
NCI CGAP\_Sub3 (IMAGE Clones 2712456-2723591) (10% of  
the driver population), plus a pool of 5,472 clones from  
NCI CGAP\_Sub4 (IMAGE Clones 2723592-2729326) (40% of the  
driver population), plus a pool of 4032 clones from  
NCI CGAP\_Sub6 (IMAGE Clones 2728969-2733190) (40% of the  
driver population). Subtraction was performed as  
previously described [Bonaldi, Lennon & Soares (1996):  
Normalization and Subtraction: Two Approaches to  
Facilitate Gene Discovery. Genome Research 6, 791-806.  
TAG\_TISSUE=brain  
TAG\_LIB=NCI CGAP\_Brn50  
TAG\_SEQ=TTTGG"

ORIGIN

Query Match	17.0%;	Score 123;	DB 2;	Length 381;	
Best Local Similarity	94.7%;	Pred. No. 5.5e-25;			
Matches 126;	Conservative	1;	Mismatches 6;	Indels 0;	Gaps 0;

QY 419 AGGTGGTGTGTCAGTCACCATAGAGATTATTAAGTCCAGGAGTAGAATCAGCCTTAC 478  
|||||  
Db 133 ATGTGGTGTGTCAGTCACCATAGAGATTATTAAGTCCAGGAGTAGAATCAGCCTTAC 74  
|||||

QY 479 AAACAGTGACCTCAACGAAGGAGATGCTGACCTGAACCCACGAAAGCGAAACTCAAAA 538  
|||||  
Db 73 AAACAGTGACCTCAACGAAGGAGATGCTGACCTGAACCCACGAGGAAACTCAAAA 14  
|||||

QY 539 GAGAAGAAAGCAA 551  
|||||  
Db 13 AAAAAAAAAAAAAA 1  
|||||

RESULT 7  
AA644590  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

AA644590  
af74a01.r1 Soares NHHMPu\_S1 Homo sapiens cDNA clone IMAGE:1047720  
5', mRNA sequence.

AA644590  
AA644590.1 GI:2569808  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 394)  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
Krieman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
Contact: Wilson RK  
Washington University School of Medicine





```
RESULT 9
AI459809/c
LOCUS
DEFINITION
  ap17h05.x1 Schiller oligodendroglioma Homo sapiens cDNA clone
IMAGE:1955673 3', mRNA sequence.
ACCESSION
  AI459809
VERSION
  AI459809.1 GI:4312690
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 358)
AUTHORS
  Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
  Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
  Martin,J., Moore,B., Scheilenberg,K., Steptoe,M., Tan,F.,
  Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  WashU-NCI Human EST Project
  Unpublished (1997)
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LML; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Seq primer: -40UP from Gibco
  High quality sequence stop: 267.
FEATURES
  source
    1..358
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:1955673"
    /sex="male"
    /tissue_type="oligodendroglioma"
    /dev_stage="44 years"
    /lab_host="SOLR"
    /clone_lib="Schiller oligodendroglioma"
    /notes="Organ: brain; Vector: pBluescript SK- (Stratagene);
    Site 1: EcoRI; Site 2: XhoI; Double-stranded cDNA was
    prepared from human oligodendroglioma using primer
    5'-GAGAGAGAGAGAGAGAGAACTAGTCTGAGT(18)-3'. An EcoRI
    adaptor was used on the 5' end of the cDNA as follows:
    5'-AATTCGACGAG-3'. The library was size-selected and
    went through one round of amplification. Average insert
    size is 1.7 kb, with a range from 0.4-12 kb. Tumor
    identification by consensus pathology; contains
    chromosome 1p and 19q deletion as determined by CGH. This
    library was constructed by Dr. Martin Schiller (Johns
    Hopkins University)."
```

```
Query Match      13.8%; Score 100; DB 1; Length 358;
Best Local Similarity 89.8%; Pred. No. 3.6e-18;
Matches 106; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCAACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 536
|||||
Db 118 AGTGTGTGTGAGTCAACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 536
|||||

QY 479 AACACTGGCCCTCAACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 536
|||||
Db 58 AACACTGGCCCTCAACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 536
|||||

RESULT 10
CB586288
LOCUS
DEFINITION
  CB586288 583 bp mRNA linear EST 03-APR-2003
  AWGNNUC:NRH5-00221-H3-A W Rat hypothalamus (10471) Rattus
  norvegicus cDNA clone nrh5-00221-h3 5', mRNA sequence.
  EST.
  Contact: Sonstegard TS
  USDA, ARS, Beltsville Agricultural Research Center
  Edlg. 200 Rm 2A, Beltsville, MD 20705, USA
  Tel: 301 504 8416
  Fax: 301 504 8414
  Email: tads@lpsl.barc.usda.gov
```

```
ACCESSION
  CB586288
VERSION
  CB586288.1 GI:29530329
KEYWORDS
  EST.
SOURCE
  Rattus norvegicus (Norway rat)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
  Rattus.
REFERENCE
  1 (bases 1 to 583)
AUTHORS
  Angen EST Program.
  Angen EST Program
  Unpublished (2003)
  Contact: Dan Fitzpatrick
  Angen, Inc
  One Angen Center Drive, Thousand Oaks, CA 91320-1799, USA
  Tel: 805 447-4881
  Plate: 00221 row: h column: 3.
  Location/Qualifiers
    1..583
    /organism="Rattus norvegicus"
    /mol_type="mRNA"
    /db_xref="taxon:10116"
    /clone="nrh5-00221-h3"
    /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; W Rat
    hypothalamus adult female Wistar rat avg. insert size 2.3
    kb fraction 6 and 7"
```

```
Query Match      13.7%; Score 98.8; DB 6; Length 583;
Best Local Similarity 72.9%; Pred. No. 9.4e-18;
Matches 124; Conservative 2; Mismatches 44; Indels 0; Gaps 0;

QY 419 AGTGTGTGTGAGTCAACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 538
|||||
Db 39 AGCGGTGTGTGAGCCACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 538
|||||

QY 479 AACACTGGCCCTCAACGAGGAGTGTGCTGACCTGACCCGAGCGAACTCAA 538
|||||
Db 99 AACCTGGTGTCTCTGAGAAAGAACCCGAGCTGAGCCGAGCTGAGAACTCAA 538
|||||

QY 539 GAGAGAAAGCAACCAAGAACTCTCTGATGCGTTTCTCAGACAAATG 588
|||||
Db 159 AAGAAAGCAACCCGAGGAGCCCTCTGATGCGTTTCTCAGACAAATG 588
|||||

RESULT 11
BE485778
LOCUS
DEFINITION
  BE485778 359 bp mRNA linear EST 27-MAR-2003
  172996 BARC 580V Bos taurus cDNA 5', mRNA sequence.
ACCESSION
  BE485778
VERSION
  BE485778.1 GI:9605311
KEYWORDS
  EST.
SOURCE
  Bos taurus (cow)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
  Bovinae; Bos.
REFERENCE
  1 (bases 1 to 359)
AUTHORS
  Sonstegard,T., Capuco,A.V., White,J., Van Tassel,C.P.,
  Connor,E.E., Cho,J., Sultana,R., Shade,L., Wray,J.E., Wells,K.D.
  and Quackenbush,J.
  Analysis of bovine mammary gland EST and functional annotation of
  the Bos taurus gene index
  Mamm. Genome 13 (7), 373-379 (2002)
  22135956
  12140684
  Contact: Sonstegard TS
  USDA, ARS, Beltsville Agricultural Research Center
  Edlg. 200 Rm 2A, Beltsville, MD 20705, USA
  Tel: 301 504 8416
  Fax: 301 504 8414
  Email: tads@lpsl.barc.usda.gov
```

Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross\_match with the -minscore 18 and -minmatch 12 options.

## PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGCAGC

Plate: 132 row: N column: 19

Seq primer: ATTAGGTGACACTATAG.

Location/Qualifiers

1. .359

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/tissue\_type="pooled"

/lab\_host="DH10B"

/clone\_lib="BARC 5BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled mRNA isolated from mammary

tissues at eight physiological, developmental, and disease

states."

states."

## ORIGIN

Query Match 12.4%; Score 89.8; DB 2; Length 359;

Best Local Similarity 76.3%; Pred. No. 3.9e-15;

Matches 135; Conservative 2; Mismatches 34; Indels 6; Gaps 2;

## QY

418 CAGGTGGTGTGAGTACACAGTAGAGATTATAAGTCCAAGGAAGTAGAATCAGCCCTTA 477

## Db

52 CAGGTGGTGTGAGTACACAGTAGAGTGTAAAGTCTGAAGAGTAGAGTCAGCCCTTA 111

## QY

478 CAACAGTGGACCT---CAACGAAGAGAGATGCTGACCTGAACCCACWAGACGGAATC 534

## Db

112 CAACAGTGGATCTCGGCGAAGAGAGAGAGCGACACCCGATCCGCGAGAAGACACCC 171

## QY

535 AAAAGAGAGAAGCAAA---CCAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588

## Db

172 AGGAGAGAAGAACGACGAGAGCCCGAGGCCCTCCCTGATGGCGCTCCTCAGACAGATG 228

## RESULT 12

## BE477506

LOCUS BE477506 . 246 bp mRNA linear EST 27-MAR-2003

DEFINITION 161238 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.

ACCESSION BE477506

VERSION BE477506.1 GI:9597039

KEYWORDS EST.

SOURCE Bos taurus (cow)

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

Bovinae; Bos.

1 (bases 1 to 246)

SONSTEGARD T., CAPUCCO A.V., WHITE J., VAN TASSELL C.P.,

CONNOR E.E., CHO J., SULTANA R., SHADE L., WRAY J.E., WELLS K.D.

and Quackenbush, J.

Analysis of bovine mammary gland EST and functional annotation of

the Bos taurus gene index

Mamm. Genome 13 (7), 373-379 (2002)

22135956

12140684

CONTACT: Sonstegard TS

USDA, ARS, Beltsville Agricultural Research Center

Bldg. 200 Rm 2A, Beltsville, MD 20705, USA

Tel: 301 504 8416

Fax: 301 504 8414

Email: tads@psi.barc.usda.gov

Single pass sequencing. Bases called and alt trimmed with phred

v0.980904.e. Vector identified by cross\_match with the -minscore 18

and -minmatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCCTCAGCAGC

Plate: 3 row: B column: 3

## FEATURES

## source

1. .246

/organism="Bos taurus"

/mol\_type="mRNA"

/db\_xref="taxon:9913"

/tissue\_type="pooled"

/lab\_host="DH10B"

/clone\_lib="BARC 5BOV"

/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;

Library made from pooled mRNA isolated from mammary

tissues at eight physiological, developmental, and disease

states."

states."

## ORIGIN

Query Match 12.2%; Score 88.2; DB 2; Length 246;

Best Local Similarity 75.7%; Pred. No. 1e-14;

Matches 134; Conservative 2; Mismatches 35; Indels 6; Gaps 2;

## QY

418 CAGGTGGTGTGAGTACACAGTAGAGATTATAAGTCCAAGGAAGTAGAATCAGCCCTTA 477

## Db

37 CAGGTGGTGTGAGTACACAGTAGAGCTGTAAAGTCTGAAGAAGTAGAGTCAGCCCTTA 96

## QY

478 CAACAGTGGACCT---CAACGAAGGAGATGCTGCACCTGAACCCACWAGACGGAATC 534

## Db

97 CAACAGTGGATCTCGGCGAAGAGAGAGAGCGACACCCGATCCGCGAGAAGACACCC 156

## QY

535 AAAAGAGAAGAAAGCAAA---CCAGAACCTCTCTGATGRCGTTTCTCAGACAAATG 588

## Db

157 GGGAGAGAGAAACGACGAGAGCCCGAGGCCCTCCCTGATGGCGCTCCTCAGACAGATG 213

## RESULT 13

## BG964319

LOCUS BG964319

DEFINITION 602831931F1 NCI\_CGAP\_Co24 Mus musculus cDNA clone IMAGE:4986798 5',

mRNA sequence.

ACCESSION BG964319

VERSION BG964319.1 GI:14351956

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 668)

NIH-MGC <http://mgc.ncl.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Jeffrey E. Green, M.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: L1AM10996 row: m column: 07

High quality sequence stop: 668.

Location/Qualifiers

1. .668

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="FVB/N"

/db\_xref="taxon:10090"

/clone="IMAGE:4986798"

/lab\_host="DH10B (T1 phage-resistant)"

/clone\_lib="NCI CGAP Co24"

/note="Organ: Colon; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.6 kb. Constructed by Life

Technologies. Note: this is a NCI\_CGAP Library."

## ORIGIN



REFERENCE  
AUTHORS

6 (bases 1 to 2339)  
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,  
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,  
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,  
Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,  
Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,  
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,  
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,  
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,  
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,  
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,  
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,  
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission  
Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,  
URL:htp://genome.gsc.riken.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)  
Please visit our web site (<http://genome.gsc.riken.jp/>) for further  
details.

TITLE  
JOURNAL

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5'-GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. Second  
strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGATTCGAGTAAATTAATATCCCTCCCCCCC 3']. cDNA was cleaved  
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.

FEATURES  
source

Host: SOLR.  
Location/Qualifiers  
1..2339  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:2210416M21"  
/db\_xref="taxon:10090"  
/clone="2210416M21"  
/sex="male"  
/tissue\_type="stomach"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
41..1413  
/note="unnamed protein product; putative  
weakly similar to BREAST CARCINOMA AMPLIFIED SEQUENCE 1  
(NOVEL AMPLIFIED IN BREAST CANCER 1) (AMPLIFIED AND  
OVEREXPRESSED IN BREAST CANCER) (Homo sapiens)  
(SWISSPROT|O75363, evidence: FASTY, 56%id, 70.3%length,  
match=1371)"  
/codon\_start=1  
/protein\_id="BAB25989.1"  
/db\_xref="GI:12843458"  
/translation="EQTLPATPPSPESQAEAPADKDFPLNRRFKLDKGRSA  
PWSQPKAESDEQATEAPVGNPHGVSAQEDIVDSRGQDVDTLSVSPGDP  
EVPYTKEDPQVVDITENSSIMSFKTLVSPNKTETKDPEDYKATKADSVCDGHA  
GQMSQAKSKKRLDPSRLGLSPKLFRRHKDTENSPTISANLKSIDKANFTPOETRG  
KTKATKSCSPPPPTSEBDSGKAGPTSLPLGLFWKSKYKEDTLSTGAENA  
VCESPVETVRLVESSSLQTVLSEETQPEPTDVKVKEKSPKPTPLMAFLRQMSVRS  
SEGI PRSESNVDSKSCOTSNSVETKTPSPPEPAGTAQKNKETSSTSSKDKKSVDKSA  
TENSQKNGKQREVPAPCVQPTVEANAMQTGDKTPKSEKRQSLQGLGPKR  
MSDAQVQIDPVSIGPVGSKS"  
2317..2322  
/note="putative"  
2339  
/note="putative"

CDS

polyA\_signal  
polyA\_site  
ORIGIN

Query Match

12.0%; Score 86.6; DB 3; Length 2339;

Best Local Similarity 60.7%; Pred. No. 5.8e-14;  
Matches 156; Conservative 2; Mismatches 96; Indels 3; Gaps 1;  
QY 332 AGGGGGGCGAGCAAGGACCCCACTTTGTAAAAACATGTTTGGGGGAATGTTTGTGTTTCA 391  
DB 737 AGGCAGGGGCCACCTCACTACCGCTGGGAAAGTTGTTTGGAAAGAAGTCAGTTAAGGAGG 796  
QY 392 TTTTCTTTATTACCTGGGCAAAATATCCAGGTGGTGTGTGAGTCCACCATAGAGATTATAA 451  
DB 797 ATACACTTTTCCACAGGTGCAGAGGAGAACGCGGTGTGTGAGTCCACCATAGAGACCGTAA 856  
QY 452 AGTCCAAGGAAGTAGAATCAAGCTTTACAAACAGTGGACCTCAACGAAGGAGATGCTGCAC 511  
DB 857 GCCTTGAGGAAGTAGAATCCAGCTTTACAAACTTGGATCTCAGTGA---AGAGACCCAGC 913  
QY 512 CTGAACCCCAWGAAGCGGAAACTCAAAAGAGAGAAAGCAAAACCAAGAACCTCTCTGTATGR 571  
DB 914 CTGAACCCCAWGAAGCGTAAAGTCAAAAGAGAAAGCAAAACCCCGGAAGACCCCTCTGTATGG 973  
QY 572 CGTTTCTCAGACAAATG 588  
DB 974 CGTTTCTCAGACAAATG 990

Search completed: May 5, 2005, 06:53:53  
Job time : 3074.62 secs

**This Page Blank (uspto)**

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:11:25 ; Search time 6286.44 Seconds  
(without alignments)  
9927.764 Million cell updates/sec

Title: US-08-731-499-5  
Perfect score: 1288  
Sequence: 1 GAGGCACGAGAGGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1288	100.0	1288	6	AR070328	AR070328 Sequence
2	1288	100.0	1288	6	BD085729	BD085729 Genes fro
3	1283.4	99.6	1830	9	HSM808295	BX648148 Homo sapi
4	1265	98.2	3322	6	AX746848	AX746848 Sequence
5	1265	98.2	3322	9	AK091206	AK091206 Homo sapi
6	1137	88.3	54161	9	BX276189	BX276189 Human DNA
7	1137	88.3	172256	2	AC051637	AC051637 Homo sapi
8	1125	87.3	57493	9	AC004505	AC004505 Homo sapi
9	1013	78.6	3168	6	C0740312	C0740312 Sequence
10	909.2	70.6	4502	10	BC079550	BC079550 Mus muscu
11	909.2	70.6	218037	10	AL731822	AL731822 Mouse DNA
12	908.2	70.5	4041	10	BC079877	BC079877 Mus muscu
13	897.2	69.7	3294	10	AF207880	AF207880 Mus muscu
14	656.4	51.0	2012	6	CQ078477	CQ078477 Sequence
15	568.6	44.1	671	11	BV014871	BV014871 S212P6046
16	527.4	40.9	1152	10	BC028776	BC028776 Mus muscu
17	521	40.5	1964	6	CQ069297	CQ069297 Sequence
18	395.2	30.7	3213	10	AF191309	AF191309 Mus muscu
19	393.6	30.6	2900	10	BC058264	BC058264 Mus muscu

20	393.6	30.6	2914	10	BC038610	BC038610 Mus muscu	
21	393.6	30.6	3215	10	BC017636	BC017636 Mus muscu	
22	393.6	30.6	4096	10	BC040215	BC040215 Mus muscu	
23	393.6	30.6	193064	10	AC124012	AC124012 Mus muscu	
c	24	393.6	30.6	252617	2	AC124231	AC124231 Mus muscu
c	25	390.4	30.3	183251	2	AC013555	AC013555 Homo sapi
26	390.4	30.3	189167	9	AC116003	AC116003 Homo sapi	
27	390.4	30.3	201027	9	AC025105	AC025105 Homo sapi	
28	389.4	30.2	2933	6	CQ715160	CQ715160 Sequence	
29	388.8	30.2	4459	9	HSW804634	AL833321 Homo sapi	
30	387.2	30.1	4854	6	CQ849970	CQ849970 Sequence	
31	387.2	30.1	4854	9	AK127042	AK127042 Homo sapi	
32	387.2	30.1	230105	2	AC108227	AC108227 Rattus no	
c	33	387.2	30.1	239857	2	AC131132	AC131132 Rattus no
34	387.2	30.1	256029	2	AC094999	AC094999 Rattus no	
c	35	387.2	30.1	312143	2	AC123509	AC123509 Rattus no
36	377.8	29.3	2918	6	AR279649	AR279649 Sequence	
37	377.8	29.3	2918	6	BD079808	BD079808 Cancer-as	
38	377.8	29.3	2918	9	AF039698	AF039698 Homo sapi	
39	283.8	22.0	3755	5	AF242292	AF242292 Danio rer	
40	283.8	22.0	145503	5	BX470238	BX470238 Zebrafish	
c	41	283.8	22.0	188687	2	BX640405	BX640405 Danio rer
42	278.4	21.6	230796	2	AC098154	AC098154 Rattus no	
c	43	276.8	21.5	110000	2	AC108426 <sup>1</sup>	Continuation (2 of
44	275.2	21.4	139125	10	AC104100	AC104100 Mus muscu	
45	275.2	21.4	179415	2	AC151001	AC151001 Mus muscu	

ALIGNMENTS

RESULT 1  
LOCUS AR070328 1288 bp DNA linear PAT 18-FEB-2000  
DEFINITION Sequence 5 from patent US 5892010.  
ACCESSION AR070328  
VERSION AR070328.1 GI:7221216  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 1288)  
AUTHORS Gray,J., Collins,C., Hwang,S.-i., Godfrey,T., Kowbel,D. and Rommens,J.  
TITLE Genes from the 20Q13 amplicon and their uses  
JOURNAL Patent: US 5892010-A 5 06-APR-1999;  
FEATURES  
source Location/Qualifiers  
1..1288  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN	Query Match	100.0%;	Score 1288;	DB 6;	Length 1288;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 1288;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Qy	1	GAGGCACGAGAGGAGAAACCCAGCCCTGGAGCCACACATCTGCTGTGAGCAATGGG	60		
Db	1	GAGGCACGAGAGGAGAAACCCAGCCCTGGAGCCACACATCTGCTGTGAGCAATGGG	60		
Qy	61	TGCGCCCTCGCCAAACACGCGCCCGCCCTGCATCATCAACCCACTCAGCGCCCTGCAG	120		
Db	61	TGCGCCCTCGCCAAACACGCGCCCGCCCTGCATCATCAACCCACTCAGCGCCCTGCAG	120		
Qy	121	TCCGTCCTGAACATCACTTGGGCAACGACGAGCCCTTGGCGCTCACTTCTGCTCTCC	180		
Db	121	TCCGTCCTGAACATCACTTGGGCAACGACGAGCCCTTGGCGCTCACTTCTGCTCTCC	180		
Qy	181	AGCCCAAGTTCAGCACAAATTTCCATGTTCCACAGTCAATCTCAATGTATGACACAAG	240		
Db	181	AGCCCAAGTTCAGCACAAATTTCCATGTTCCACAGTCAATCTCAATGTATGACACAAG	240		
Qy	241	CCGGTCTTGGTCTGCTGCCCTCCCAAGGTTCAGCCAGCGTGTCCAGGCGGTACCTGTTGAG	300		

|||||  
241 CCGGTCTTGAGTCTGCTCCACAGGTGAGCCAGCGTGTCCAGGCGCTACCTGTTTGAG 300  
|||||  
301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360  
|||||  
301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360  
|||||  
361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCCGACATGGTTC 420  
|||||  
361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCCGACATGGTTC 420  
|||||  
421 AAAGTCTCTCCAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGTTCCCGCCCATG 480  
|||||  
421 AAAGTCTCTCCAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGTTCCCGCCCATG 480  
|||||  
481 AAGCTGGAATGATGTCTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540  
|||||  
481 AAGCTGGAATGATGTCTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540  
|||||  
541 AAAAGAAAAGGCGCGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCTACAAGCCCAAG 600  
|||||  
541 AAAAGAAAAGGCGCGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCTACAAGCCCAAG 600  
|||||  
601 TTGCTCTCAGGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660  
|||||  
601 TTGCTCTCAGGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660  
|||||  
661 CAAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720  
|||||  
661 CAAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720  
|||||  
721 CTGGCCAAAGTCAAGTACAGCTTTAGGAAAACGGGGGACAAAAATTTCTGAAAAACATG 780  
|||||  
721 CTGGCCAAAGTCAAGTACAGCTTTAGGAAAACGGGGGACAAAAATTTCTGAAAAACATG 780  
|||||  
781 GACAAAGGCCACCCCATCTTTTATTGAGTGAATCTGCTCCAGTTTCAGAAACCCCTTCT 840  
|||||  
781 GACAAAGGCCACCCCATCTTTTATTGAGTGAATCTGCTCCAGTTTCAGAAACCCCTTCT 840  
|||||  
841 ACCTACATCAGTCACTTAGATCTCAGCTGGGTTTCCAAATGAGACATGACCCGCTTG 900  
|||||  
841 ACCTACATCAGTCACTTAGATCTCAGCTGGGTTTCCAAATGAGACATGACCCGCTTG 900  
|||||  
901 TCAGTGGACCAAGCAAGAGGTGGAGCAAGAGATCTCCCGGGTATCTCGGCTCAGAGG 960  
|||||  
901 TCAGTGGACCAAGCAAGAGGTGGAGCAAGAGATCTCCCGGGTATCTCGGCTCAGAGG 960  
|||||  
961 TCTCCAGAAAACATAGCTGCGGAAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020  
|||||  
961 TCTCCAGAAAACATAGCTGCGGAAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020  
|||||  
1021 TGTGCGACATTTGTGAGCAACATGCGGTAAACTCCACCTTANGCAAAAGCCACAGCAAG 1080  
|||||  
1021 TGTGCGACATTTGTGAGCAACATGCGGTAAACTCCACCTTANGCAAAAGCCACAGCAAG 1080  
|||||  
1081 TCACCCGAACACCATTTACAGTTTGTAAACAGAGCTGATGAAGAAATAGCTCTCAGGACG 1140  
|||||  
1081 TCACCCGAACACCATTTACAGTTTGTAAACAGAGCTGATGAAGAAATAGCTCTCAGGACG 1140  
|||||  
1141 AATGSCCTTAGTTCCTACTTTCCAGCCTGGATCCCTCACACTGAACCCCTTTCTCGTTGCA 1200  
|||||  
1141 AATGSCCTTAGTTCCTACTTTCCAGCCTGGATCCCTCACACTGAACCCCTTTCTCGTTGCA 1200  
|||||  
1201 CGATCTGCTCTGACATTTGAATCTGATTAAGTCTCTGAGACCTGGTCTGAGAAGA 1260  
|||||  
1201 CGATCTGCTCTGACATTTGAATCTGATTAAGTCTCTGAGACCTGGTCTGAGAAGA 1260  
|||||  
1261 CTGCCAAAAAATAAAAAAAAAAATTC 1288  
|||||  
1261 CTGCCAAAAAATAAAAAAAAAAATTC 1288

BD085729 1288 bp DNA linear PAT 27-AUG-2002  
LOCUS Genes from the 20q13 amplicon and their uses.  
DEFINITION  
ACCESSION BD085729  
VERSION BD085729.1 GI:22631339  
KEYWORDS JP 2001524802-A/5.  
SOURCE synthetic construct  
ORGANISM synthetic construct  
REFERENCE  
1 (bases 1 to 1288)  
AUTHORS Gray,J.W., Collins,C.C., Hwang,S.I., Godfrey,T., Kowbel,D. and Rommens,J.  
TITLE Genes from the 20q13 amplicon and their uses  
JOURNAL Patent: JP 2001524802-A 5 04-DEC-2001;  
COMMENT THE REGENTS OF THE UNIVERSITY OF CALIFORNIA  
OS Artificial Sequence  
PN JP 2001524802-A/5  
PD 04-DEC-2001  
PF 15-JUL-1997 JP 1998506264  
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR  
17-JAN-1997 US 08/785532  
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI DAVID KOWBEL,  
PI JOHANNA ROMMENS  
PC C12N15/11, C12Q1/68, A61K48/00  
CC Description of Artificial Sequence:41.1  
FH Key Location/Qualifiers  
FT source 1..1288  
FT /organism='Artificial Sequence'.  
FEATURES  
source  
1..1288  
/organism='synthetic construct'  
/mol\_type='genomic DNA'  
/db\_xref='taxon:32630'  
ORIGIN  
Query Match 100.0%; Score 1288; DB 6; Length 1288;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAGGGCAGCGAAGAGAGAAACCCAGAGCCCTTGGAGCCCAATCTGCTCTGAGCAATGGG 60  
DB 1 GAGGGCAGCGAAGAGAGAAACCCAGAGCCCTTGGAGCCCAATCTGCTCTGAGCAATGGG 60  
QY 61 TCGGCTCTCGCAACCAAGCAGCCCGCTTGCATGCATCAACCCACTCAGCGCCCTGCAAG 120  
DB 61 TCGGCTCTCGCAACCAAGCAGCCCGCTTGCATGCATCAACCCACTCAGCGCCCTGCAAG 120  
QY 121 TCGGCTCTCGCAACCAATCACTTGGGCAAGCCAGGAGCCCTTGCCTTCACTTCTGCTCC 180  
DB 121 TCGGCTCTCGCAACCAATCACTTGGGCAAGCCAGGAGCCCTTGCCTTCACTTCTGCTCC 180  
QY 181 AGCCCAAGTTCAAGCACAATTTCCATGTTTCCACAAGTCAATCTCAATGTCATGCAAG 240  
DB 181 AGCCCAAGTTCAAGCACAATTTCCATGTTTCCACAAGTCAATCTCAATGTCATGCAAG 240  
QY 241 CCGGCTCTGAGTCTCGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGG 300  
DB 241 CCGGCTCTGAGTCTCGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGG 300  
QY 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360  
DB 301 AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCAA 360  
QY 361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCCGACATGGTTC 420  
DB 361 GCACAAATCTTGATGTCTCCACCTCAGAAAGCAGCTCTGTCTGACATCGCCGACATGGTTC 420  
QY 421 AAAGTCTCTCCAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGTTCCCGCCCATG 480  
DB 421 AAAGTCTCTCCAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGTTCCCGCCCATG 480  
QY 481 AAGCTGGAATGATGTCTCAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540



```
Db 481 AAGCTGGAATGGATGTGAGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGAT 540
Qy 541 AAAAGAAAAGCGCGCAGTCAACTGGAATCTCAGCATCTTCTGATTTCAAGCCCGAG 600
Db 541 AAAAGAAAAGCGCGCAGTCCAACTGGAATCTCAGCATCTTCTGATTTCAAGCCCGAG 600
Qy 601 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660
Db 601 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660
Qy 661 CAAGAGCGTATGCAATCTCTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 661 CAAGAGCGTATGCAATCTCTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720
Qy 721 CTGGCCAAAGCTCACTAGTACAGCTTAGGAAAACCGCGGGGACAAAATTTCTGAAAACATG 780
Db 721 CTGGCCAAAGCTCACTAGTACAGCTTAGGAAAACCGCGGGGACAAAATTTCTGAAAACATG 780
Qy 781 GACAAAGGCCACCCCATCTTTTATTCAGTGTGCTGCTCCAGTTGAGAACCCCTTCT 840
Db 781 GACAAAGGCCACCCCATCTTTTATTCAGTGTGCTGCTCCAGTTGAGAACCCCTTCT 840
Qy 841 ACCTACATCAGTCACTAGTACCTCACTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Db 841 ACCTACATCAGTCACTAGTACCTCACTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Qy 901 TCAGTGGACCAAGCAAGGTCAGGCAAGAGATCTCCCGGTTATCGTGGCTCAGAGG 960
Db 901 TCAGTGGACCAAGCAAGGTCAGGCAAGAGATCTCCCGGTTATCGTGGCTCAGAGG 960
Qy 961 TCTCCAGAAAACATAGCTGCGGAGGACACAGACTCTAAATTCAGTGAAGTTGTG 1020
Db 961 TCTCCAGAAAACATAGCTGCGGAGGACACAGACTCTAAATTCAGTGAAGTTGTG 1020
Qy 1021 TGTCCGACATTTGTGAGCAACATCGGTAAATCTCCACCTTAAGCAAAACGACAGCAAG 1080
Db 1021 TGTCCGACATTTGTGAGCAACATCGGTAAATCTCCACCTTAAGCAAAACGACAGCAAG 1080
Qy 1081 TCACCCGAAACACATTCAGTGTGTAACAGACGTGGATGAAGATAGCTCTGACGAGAG 1140
Db 1081 TCACCCGAAACACATTCAGTGTGTAACAGACGTGGATGAAGATAGCTCTGACGAGAG 1140
Qy 1141 AATGCTTAGTTTCCACTTTCCAGCTGGATCCCTCAGCTGAACCCCTTCTGCTGCA 1200
Db 1141 AATGCTTAGTTTCCACTTTCCAGCTGGATCCCTCAGCTGAACCCCTTCTGCTGCA 1200
Qy 1201 CCAATCTGCTTCTGACATTTGAATCTCAATTTGAACTCTCTGACACCCCTTGCTGAGAAGA 1260
Db 1201 CCAATCTGCTTCTGACATTTGAATCTCAATTTGAACTCTCTGACACCCCTTGCTGAGAAGA 1260
Qy 1261 CTGCCAAAAAATAAAAAAAAAAATTC 1288
Db 1261 CTGCCAAAAAATAAAAAAAAAAATTC 1288

RESULT 3
HSM808295 1830 bp mRNA linear PRI 30-AUG-2003
LOCUS Homo sapiens mRNA; cDNA DKFZp686K2480 (from clone DKFZp686K2480).
DEFINITION BX648148
ACCESSION BX648148.1 GI:34367307
VERSION 1
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1830)
AUTHORS Ansoorge W., Krieger, S., Regiert, T., Rittmüller, C., Schwager, B.,
Mewes, H.W., Weill, B., Amd, C., Osanger, A., Fobo, G., Han, M. and
Wiemann, S.
CONSRMT The German Human cDNA Consortium
TITLE Direct Submission
```

```
JOURNAL Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/EMBL (European Molecular Biology Laboratories,
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp686K2480) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
Location/Qualifiers
source
1..1830
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686K2480"
/tissue_type="human endometrium carcinoma cell line"
/clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
DH10B; sites SfiIA + SfiIB"
/dev_stage="adult"
1819
polyA_site
ORIGIN
Query Match 99.6%; Score 1283.4; DB 9; Length 1830;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1284; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 GAGGGCAGGAGAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60
Db 300 GAGGGCAGGAGAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 359
Qy 61 TGGCGCTTCGCAACACACGCGCCCTGCGCATCATCAACCCACTCAGGCGCTGCGAG 120
Db 360 TGGCGCTTCGCAACACACGCGCCCTGCGCATCATCAACCCACTCAGGCGCTGCGAG 419
Qy 121 TCGGTCTGAAACAATCACTTGGGCAAGCAACGAGCCCTTGGCGTCACTTCTGCTCC 180
Db 420 TCGGTCTGAAACAATCACTTGGGCAAGCAACGAGCCCTTGGCGTCACTTCTGCTCC 479
Qy 181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAAGTCGAATCTCAATGTCAAGCAAG 240
Db 480 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAAGTCGAATCTCAATGTCAAGCAAG 539
Qy 241 CCGGTCTTGAAGTCTGCTCCACAAGGTCAGCAGCGGTGTCAGGCGGTACCTGTTTGAG 300
Db 540 CCGGTCTTGAAGTCTGCTCCACAAGGTCAGCAGCGGTGTCAGGCGGTACCTGTTTGAG 599
Qy 301 AACAGCGATCAGCCCAATTTGACTGACCAAGTCCAAAAGCAAGAGCGAGTCTCGCA 360
Db 600 AACAGCGATCAGCCCAATTTGACTGACCAAGTCCAAAAGCAAGAGCGAGTCTCGCA 659
Qy 361 GCACATCTTGTATGTCCTCCACCTCAGAGCAGCGCTGTCGACATCGCGACATGGTC 420
Db 660 GCACATCTTGTATGTCCTCCACCTCAGAGCAGCGCTGTCGACATCGCGACATGGTC 719
Qy 421 AAGTCTCTCCCAAGCCACCAACCCCAAGCAGCGCTCTCTCTCCAGGGTCCCCCCCCATG 480
Db 720 AAGTCTCTCCCAAGCCACCAACCCCAAGCAGCGCTCTCTCTCCAGGGTCCCCCCCCATG 779
Qy 481 AAGCTGGAATTTGAGTGTGAGGCGCTTTGAGGATGCTCAAGTGAAGTCTCAACTTTGCA 540
Db 780 AAGCTGGAATTTGAGTGTGAGGCGCTTTGAGGATGCTCAAGTGAAGTCTCAACTTTGCA 839
Qy 541 AAAAGAAAAGCGCGCAGTCCAACTGGAATCTCAGCATCTTCTGATTTCAAGCCCGAG 600
Db 840 AAAAGAAAAGCGCGCAGTCCAACTGGAATCTCAGCATCTTCTGATTTCAAGCCCGAG 899
Qy 601 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660
Db 900 TTTGCTCGAGCTTTCCAGACATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCA 959
```

```
QY 661 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACCTATCAGTCACTGG 720
Db 960 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACCTATCAGTCACTGG 1019
QY 721 CTGGCCAAAGTCAAGTACAGCTTAGAATAACGGGGGGGACAAAATTTCTGAAAAACATG 780
Db 1020 CTGGCCAAAGTCAAGTACAGCTTAGAATAACGGGGGGGACAAAATTTCTGAAAAACATG 1079
QY 781 GACAAAGGCCACCCCTCTTTTATTCAGTGACTGTGCCCTCCAGTTTCAGAACCCCTTCT 840
Db 1080 GACAAAGGCCACCCCTCTTTTATTCAGTGACTGTGCCCTCCAGTTTCAGAACCCCTTCT 1139
QY 841 ACCTACATCAGTCACCTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTG 900
Db 1140 ACCTACATCAGTCACCTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTG 1199
QY 901 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 960
Db 1200 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 1259
QY 961 TCTCCAGAAACAATAGCTGCGGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 1020
Db 1260 TCTCCAGAAACAATAGCTGCGGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 1319
QY 1021 TGTCCGACATTTGTGAGCAAAACATCGCGTAAAACTCCACTAAGCAAAAACGCACAGCAAG 1080
Db 1320 TGTCCGACATTTGTGAGCAAAACATCGCGTAAAACTCCACTAAGCAAAAACGCACAGCAAG 1379
QY 1081 TCACCCGAACACATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGCTTCGAGGAG 1140
Db 1380 TCACCCGAACACATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGCTTCGAGGAG 1439
QY 1141 AATGCTTGTAGTTTCCACTTTCAGAGCTGGATCCCTCAGACTGAAACCTTCTCGTTGCA 1200
Db 1440 AATGCTTGTAGTTTCCACTTTCAGAGCTGGATCCCTCAGACTGAAACCTTCTCGTTGCA 1499
QY 1201 CCATCTGCTTTCAGATTAATTCATTAATTCCTCCTGACACCTGGCTCTGAGAAGA 1260
Db 1500 CCATCTGCTTTCAGATTAATTCATTAATTCCTCCTGACACCTGGCTCTGAGAAGA 1559
QY 1261 CTGCCAAAAAATAAAAAAAAAA 1285
Db 1560 CTGCCAAAAAATAAAAAAAAAA 1584

RESULT 4
AX746848
LOCUS AX746848
DEFINITION Sequence 373 from Patent EP1308459.
ACCESSION AX746848
VERSION AX746848.1 GI:32131236
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
Masuho,Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1308459-A 373 07-MAY-2003;
Helix Research Institute (JP) ; Research Association for
Biotechnology (JP)
FEATURES
source
1..3322
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

Query Match 98.2%; Score 1265; DB 6; Length 3322;
```

```
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGGGCAGCGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60
Db 2058 GAGGGCAGCGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 2117
QY 61 TCGCCCTCGCCAAACACAGCCCGCCCTGCCATGTCATCAACCCACTCAGCGCCCTGCAG 120
Db 2118 TCGCCCTCGCCAAACACAGCCCGCCCTGCCATGTCATCAACCCACTCAGCGCCCTGCAG 2177
QY 121 TCCGTCCTGAAACAATCATTGGGCAAGCCAGCGCCCTTCCGCTCACCCTTCTGCTCC 180
Db 2178 TCCGTCCTGAAACAATCATTGGGCAAGCCAGCGCCCTTCCGCTCACCCTTCTGCTCC 2237
QY 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAG 240
Db 2238 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAG 2297
QY 241 CCGGTCCTGAGTCTGCCCTCCACAAGGTCAGCCAGCGCTCCAGGCGCTACCTGTTTGAG 300
Db 2298 CCGGTCCTGAGTCTGCCCTCCACAAGGTCAGCCAGCGCTCCAGGCGCTACCTGTTTGAG 2357
QY 301 AACAGCGATFAGCCCAATTGACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCTCGCAA 360
Db 2358 AACAGCGATFAGCCCAATTGACCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCTCTCGCAA 2417
QY 361 GCACAATCTTGATGTGCCCACTCAGAAGCAGCTGTCTGACATCGCCGACATGGTC 420
Db 2418 GCACAATCTTGATGTGCCCACTCAGAAGCAGCTGTCTGACATCGCCGACATGGTC 2477
QY 421 AAAGTCTCTCCCAAGCCACACCCCAAGCCAGCTCTCTCCAGGGTCCCCCCCATG 480
Db 2478 AAAGTCTCTCCCAAGCCACACCCCAAGCCAGCTCTCTCTCCAGGGTCCCCCCCATG 2537
QY 481 AAGCTGGAAATGGATGTTCAGGGCTTTGAGGATGTCTCCAGTCAAGTCTCAACTTTGCAT 540
Db 2538 AAGCTGGAAATGGATGTTCAGGGCTTTGAGGATGTCTCCAGTCAAGTCTCAACTTTGCAT 2597
QY 541 AAAAGAAAAGGGCGGAGTCCAACTGGGAATCTCTCAGCATCTTGTGATCTACAGGCCAG 600
Db 2598 AAAAGAAAAGGGCGGAGTCCAACTGGGAATCTCTCAGCATCTTGTGATCTACAGGCCAG 2657
QY 601 TTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTCATCTGGGCCCA 660
Db 2658 TTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTCATCTGGGCCCA 2717
QY 661 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 2718 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 2777
QY 721 CTGGCCAAAGTCAAGTACCAGCTTAGAATAACGGGGGGGACAAAATTTCTGAAAAACATG 780
Db 2778 CTGGCCAAAGTCAAGTACCAGCTTAGAATAACGGGGGGGACAAAATTTCTGAAAAACATG 2837
QY 781 GACAAAGGCCACCCCTCTTTTATTCAGTGACTGTGCCCTCCAGTTTCAGAACCCCTTCT 840
Db 2838 GACAAAGGCCACCCCTCTTTTATTCAGTGACTGTGCCCTCCAGTTTCAGAACCCCTTCT 2897
QY 841 ACCTACATCAGTCACCTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTG 900
Db 2898 ACCTACATCAGTCACCTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCGCTTG 2957
QY 901 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 960
Db 2958 TCAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGAGG 3017
QY 961 TCTCCAGAAACAATAGCTGCGGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 1020
Db 3018 TCTCCAGAAACAATAGCTGCGGAAGAGGACACAGACTCTAAATTCAGGTGAAGTTGTGC 3077
QY 1021 TCTCCGACATTTGTGAGCAAAACATCGCGTAAATCTCCACTAAGCAAAAACGCACAGCAAG 1080
Db 1080 TCTCCGACATTTGTGAGCAAAACATCGCGTAAATCTCCACTAAGCAAAAACGCACAGCAAG 1080
```



Db 2298 CCGGTCTTGAGTCTCGCTCCACAAAGGTGAGCCAGCGTGTCCAGGCGCTACCTGTTTGAG 2357  
Qy 301 AACAGGATCAGCCCATTCACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCGCA 360  
Db 2358 AACAGGATCAGCCCATTCACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCGCA 2417  
Qy 361 GCACAAATCTTGATGTGCTCCACCTCAGAAAGCAGCTGTCTGTGAATCGCCGACATGGTC 420  
Db 2418 GCACAAATCTTGATGTGCTCCACCTCAGAAAGCAGCTGTCTGTGAATCGCCGACATGGTC 2477  
Qy 421 AAAGTCTCTCCCAAGACCAACCCCAAGCAGGCTCTCTCTCCAGGTCCTCCGCCCATG 480  
Db 2478 AAAGTCTCTCCCAAGACCAACCCCAAGCAGGCTCTCTCTCCAGGTCCTCCGCCCATG 2537  
Qy 481 AAGCTGGAATGGATGTGAGCGCTTGAAGGATGTCTCCAGTGAAGTCTCAACTTGGAT 540  
Db 2538 AAGCTGGAATGGATGTGAGCGCTTGAAGGATGTCTCCAGTGAAGTCTCAACTTGGAT 2597  
Qy 541 AAAAGAAAAGGCGGCGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTACAAGCCCA 600  
Db 2598 AAAAGAAAAGGCGGCGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTACAAGCCCA 2657  
Qy 601 TTGCTCTCAGCTCTTCCAGACATCAGAGGCAAAATACCTGTGTCTGATCTGGGCCCA 660  
Db 2658 TTGCTCTCAGCTCTTCCAGACATCAGAGGCAAAATACCTGTGTCTGATCTGGGCCCA 2717  
Qy 661 CAAGAGCTATGCAATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720  
Db 2718 CAAGAGCTATGCAATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 2777  
Qy 721 CTGGCCAAAGTCAAGTACCAGCTTAGGAAAACGGGGGACAAAATTTCTGAAAACATG 780  
Db 2778 CTGGCCAAAGTCAAGTACCAGCTTAGGAAAACGGGGGACAAAATTTCTGAAAACATG 2837  
Qy 781 GACAAAGGCCACCCATCTTTTATTCAGTGAATGTGCTCCAGTTCAGAAACCCCTTCT 840  
Db 2838 GACAAAGGCCACCCATCTTTTATTCAGTGAATGTGCTCCAGTTCAGAAACCCCTTCT 2897  
Qy 841 ACCTACATCAGTCACTTAGAATCTACCTGGGTTTCCAAATGAAGACATGACCGCTTG 900  
Db 2898 ACCTACATCAGTCACTTAGAATCTACCTGGGTTTCCAAATGAAGACATGACCGCTTG 2957  
Qy 901 TCAGTGGACAGCAAGCAAGTGGAGCAAGAGATCTCCCGGTATCTCGTGGTTCAGAGG 960  
Db 2958 TCAGTGGACAGCAAGCAAGTGGAGCAAGAGATCTCCCGGTATCTCGTGGTTCAGAGG 3017  
Qy 961 TCTCCAGAAACAAATAGCTGCGGAAGAGGACACAGACTCTAAATTTCAAGTGTAAAGTGTGC 1020  
Db 3018 TCTCCAGAAACAAATAGCTGCGGAAGAGGACACAGACTCTAAATTTCAAGTGTAAAGTGTGC 3077  
Qy 1021 TGTGCGACATTTGTGAGCAAAATCGGTAAACTCCACTAAGCAAAACGCACAGCAAG 1080  
Db 3078 TGTGCGACATTTGTGAGCAAAATCGGTAAACTCCACTAAGCAAAACGCACAGCAAG 3137  
Qy 1081 TCACCCGAACACCAATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGTCTTCGAGGAG 1140  
Db 3138 TCACCCGAACACCAATTCACAGTTTGTAAACAGAGCTGGATGAAGATAGTCTTCGAGGAG 3197  
Qy 1141 AATGCTTATGTTTCCATTTTCCAGCTTGAATCCCTCACAAGTGAACCCCTTCTGTTTGA 1200  
Db 3198 AATGCTTATGTTTCCATTTTCCAGCTTGAATCCCTCACAAGTGAACCCCTTCTGTTTGA 3257  
Qy 1201 CGATCTGCTTGAATGAACTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCTTGAATGA 1260  
Db 3258 CGATCTGCTTGAATGAACTTGAATCTTGAATCTTGAATCTTGAATCTTGAATCTTGAATGA 3317  
Qy 1261 CTGCC 1265  
Db 3318 CTGCC 3322

RESULT 6  
BX276189

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BX276189  
Human DNA sequence from clone RP4-678D15 on chromosome  
20q13.13-13.2, complete sequence.  
BX276189  
BX276189.1 GI:28564238  
HTG.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 54161)  
Direct Submission  
Matthews, L.  
Submitted (25-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
Center: Wellcome Trust Sanger Institute  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest, except on the rare  
occasion of the clone being a YAC.  
The following abbreviations are used to associate primary accession  
numbers given in the feature table with their source databases:  
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information  
on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 20, constructed by the Sanger Centre Chromosome 20  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr20  
RP4-678D15 is from the library RPCI-4 constructed by the group of  
Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pCYPAC2.

FEATURES  
Location/Qualifiers

1..54161  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="Q13.13-13.2"  
/clone="RP4-678D15"  
/clone\_lib="RPCI-4"

ORIGIN

Query Match 88.3%; Score 1137; DB 9; Length 54161;  
Best Local Similarity 100.0%; Pred. No. 8e-310;  
Matches 1137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GAGGGCAGCGAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTGAGCAATGGG 60  
Db 40512 GAGGGCAGCGAGAGGAGAAACCCAGCCCTGGAGCCACATCTCTGAGCAATGGG 40571  
Qy 61 TGGCCCTCGCAACCCAGCCCTGGCCATGATCAACCCACTCAGCGCCCTGCAG 120  
Db 40572 TGGCCCTCGCAACCCAGCCCTGGCCCTGATGATCAACCCACTCAGCGCCCTGCAG 40631  
Qy 121 TCGTCTCTGAACAATCATTGGGCAAGCCAGGAGCCCTTGGCTCACTCTCTCTCTC 180

```
Db 40632 TCCGTCCTGAAACAATCACTTGGGCAAGACCGAGGCCCTTGGCGTCACTTCTGCTCC 40691
QY 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTTCCACAAGTCCAATCTCAATGTCTATGACAAAG 240
Db 40692 AGCCCAAGTTCAAGCACAAATTTCCATGTTTCCACAAGTCCAATCTCAATGTCTATGACAAAG 40751
QY 241 CCGGTCCTTGAAGTCTGCTCCCTCCACAAGGTCAGCAGCGTGTCCAGCGCTTACCTGTTTGAG 300
Db 40752 CCGGTCCTTGAAGTCTGCTCCCTCCACAAGGTCAGCAGCGTGTCCAGCGCTTACCTGTTTGAG 40811
QY 301 AACAGCGATCAGCCCAATGACTGACCAAGTCCAAAGCAAGCAAGCGAGTCTCTCGCAA 360
Db 40812 AACAGCGATCAGCCCAATGACTGACCAAGTCCAAAGCAAGCAAGCGAGTCTCTCGCAA 40871
QY 361 GCACAATCTTGTATGTCCTCCCACTCAGAGCACGCTCTGTCTGTGACATCGCGGACATGTC 420
Db 40872 GCACAATCTTGTATGTCCTCCCACTCAGAGCACGCTCTGTCTGTGACATCGCGGACATGTC 40931
QY 421 AAAGTCCTCCCAAGCACACACCCCAAGGCGCTCTCTCTCCAGGGTCCCGCCCATG 480
Db 40932 AAAGTCCTCCCAAGCACACACCCCAAGGCGCTCTCTCTCCAGGGTCCCGCCCATG 40991
QY 481 AAGCTGGAATGGATGTCAGGCGCTTTCAGGATGTCCTCAGTGAAGTCTCAACTTTCAT 540
Db 40992 AAGCTGGAATGGATGTCAGGCGCTTTCAGGATGTCCTCAGTGAAGTCTCAACTTTCAT 41051
QY 541 AAAAGAAAAGCGCGCAGTCAACTGGAATCTCAGCATCTTCTGATTTCTCAAGACCCAG 600
Db 41052 AAAAGAAAAGCGCGCAGTCAACTGGAATCTCAGCATCTTCTGATTTCTCAAGACCCAG 41111
QY 601 TTTGCTCTCAGCGCTTTCAGACATCAGAGGCAAACTCTGCTGTCTGATCTCGGGCCA 660
Db 41112 TTTGCTCTCAGCGCTTTCAGACATCAGAGGCAAACTCTGCTGTCTGATCTCGGGCCA 41171
QY 661 CAAGAGCGTATGCAATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 41172 CAAGAGCGTATGCAATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGG 41231
QY 721 CTGGCAACGTCAGTACGCTTAGGAAACGGCGGACAAATTTCTGAAAACATG 780
Db 41232 CTGGCAACGTCAGTACGCTTAGGAAACGGCGGACAAATTTCTGAAAACATG 41291
QY 781 GACAAAGCCACCCCATCTTTATTCAGTGAAGTGTGCTCTCCAGTTCAGAACCCCTTCT 840
Db 41292 GACAAAGCCACCCCATCTTTATTCAGTGAAGTGTGCTCTCCAGTTCAGAACCCCTTCT 41351
QY 841 ACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCGCTTG 900
Db 41352 ACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCGCTTG 41411
QY 901 TCAGTGGACCAAGCAAGTGGAGCAAGAGATCTCCCGGTTATCTCGGCTCAGAGG 960
Db 41412 TCAGTGGACCAAGCAAGTGGAGCAAGAGATCTCCCGGTTATCTCGGCTCAGAGG 41471
QY 961 TCTCAGAAACAATAGCTGCGGAGAGGACACAGACTCTAAATCTCAAGTGAAGTTGTC 1020
Db 41472 TCTCAGAAACAATAGCTGCGGAGAGGACACAGACTCTAAATCTCAAGTGAAGTTGTC 41531
QY 1021 TGTGGCAATTTGTGAGCAAAATCGGTAAATCTCACTTAAAGCAAAACGACGACGAG 1080
Db 41532 TGTGGCAATTTGTGAGCAAAATCGGTAAATCTCACTTAAAGCAAAACGACGACGAG 41591
QY 1081 TCACCCGAACACCATTTACAGTTTCTTAACAGAGTGGATGAAGATAGCTCTGCGAGG 1137
Db 41592 TCACCCGAACACCATTTACAGTTTCTTAACAGAGTGGATGAAGATAGCTCTGCGAGG 41648
```

RESULT 7

AC051637/c

LOCUS

DEFINITION

Homo sapiens chromosome 20 clone RP11-664G18 map 20, WORKING DRAFT

SEQUENCE, 27 unordered pieces.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

TITLE  
JOURNAL  
COMMENT

AC051637  
AC051637.2 GI:7770475  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 172556)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 20, clone RP11-664G18  
Unpublished

2 (bases 1 to 172556)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campomione,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pflieger,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Teafaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (15-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 172556)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campomione,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Gage,D.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Pflieger,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Teafaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 12, 2000 this sequence version replaced gi:7574851.  
All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L6891

Center clone name: 664\_G.18

----- Summary Statistics -----  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 155286 bases at least Q40  
Consensus quality: 164921 bases at least Q30  
Consensus quality: 168079 bases at least Q20  
Insert size: 176000; agarose-fp  
Insert size: 169656; sum-of-contigs  
Quality coverage: 3.9 in Q20 bases; agarose-fp  
Quality coverage: 4.0 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 27 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 1099: contig of 1099 bp in length  
\* 1100 1199: gap of 100 bp  
\* 1200 3175: contig of 1976 bp in length  
\* 3176 3275: gap of 100 bp  
\* 3276 3416: contig of 141 bp in length  
\* 3417 3516: gap of 100 bp  
\* 3517 4641: contig of 1125 bp in length  
\* 4642 4741: gap of 100 bp  
\* 4742 6561: contig of 1820 bp in length  
\* 6562 6661: gap of 100 bp  
\* 6662 8188: contig of 1527 bp in length  
\* 8189 8288: gap of 100 bp  
\* 8289 9545: contig of 1257 bp in length  
\* 9546 9645: gap of 100 bp  
\* 9646 11418: contig of 1773 bp in length  
\* 11419 11518: gap of 100 bp  
\* 11519 14803: contig of 3285 bp in length  
\* 14804 14903: gap of 100 bp  
\* 14904 17639: contig of 2736 bp in length  
\* 17640 17739: gap of 100 bp  
\* 17740 22510: contig of 4771 bp in length  
\* 22511 22610: gap of 100 bp  
\* 22611 27218: contig of 4608 bp in length  
\* 27219 27318: gap of 100 bp  
\* 27319 30622: contig of 3304 bp in length  
\* 30623 30722: gap of 100 bp  
\* 30723 33882: contig of 3160 bp in length  
\* 33883 33982: gap of 100 bp  
\* 33983 37898: contig of 3916 bp in length  
\* 37899 41116: contig of 3118 bp in length  
\* 41117 41216: gap of 100 bp  
\* 41217 47018: contig of 5802 bp in length  
\* 47019 47118: gap of 100 bp  
\* 47119 53665: contig of 6547 bp in length  
\* 53666 53765: gap of 100 bp  
\* 53766 60112: contig of 6347 bp in length  
\* 60113 60212: gap of 100 bp  
\* 60213 70207: contig of 9995 bp in length  
\* 70208 70307: gap of 100 bp  
\* 70308 78838: contig of 8531 bp in length  
\* 78839 78938: gap of 100 bp  
\* 78939 87552: contig of 8614 bp in length  
\* 87553 87652: gap of 100 bp  
\* 87653 104330: contig of 16678 bp in length  
\* 104331 104430: gap of 100 bp  
\* 104431 122770: contig of 18340 bp in length  
\* 122771 122870: gap of 100 bp  
\* 122871 137929: contig of 15059 bp in length  
\* 137930 138029: gap of 100 bp  
\* 138030 152291: contig of 14262 bp in length  
\* 152292 152391: gap of 100 bp  
\* 152392 172256: contig of 19865 bp in length.

FEATURES		Location/Qualifiers
misc_feature	source	1. 172256 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="20" /map="20" /clone="RP11-664G18" /clone_lib="RPC1-11 Human Male BAC"
		1. 1099 /note="assembly_fragment"
		1200. 3175 /note="assembly_fragment"
		3276. 3416 /note="assembly_fragment"
		clone_end:SP6 vector_side:right"
		3517. 4641 /note="assembly_fragment"
		4742. 6561 /note="assembly_fragment"
		6662. 8188 /note="assembly_fragment"
		8289. 9545 /note="assembly_fragment"
		9646. 11418 /note="assembly_fragment"
misc_feature		11519. 14803 /note="assembly_fragment"
		14904. 17639 /note="assembly_fragment"
		17740. 22510 /note="assembly_fragment"
		22611. 27218 /note="assembly_fragment"
		27319. 30622 /note="assembly_fragment"
		30723. 33882 /note="assembly_fragment"
		33983. 37898 /note="assembly_fragment"
		clone_end:P7 vector_side:right"
		37999. 41116 /note="assembly_fragment"
		41217. 47018 /note="assembly_fragment"
misc_feature		47119. 53665 /note="assembly_fragment"
		53766. 60112 /note="assembly_fragment"
		60212. 70207 /note="assembly_fragment"
		70307. 78838 /note="assembly_fragment"
		78839. 87552 /note="assembly_fragment"
		87553. 104330 /note="assembly_fragment"
		104331. 122770 /note="assembly_fragment"
		122771. 137929 /note="assembly_fragment"
		137930. 152291 /note="assembly_fragment"
		152292. 172256 /note="assembly_fragment"
Query Match		88.3%; Score 1137; DB 2; Length 172256;
Best Local Similarity		100.0%; Pred. No. 9.3e-310;
Matches 1137; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
Qy	1	GAGGGCAGCGAGAAGGAGAAACCCAGCCCTGGAGCCACACATCTGCTCTGAGCAATGGG 60 
Db	50899	GAGGGCAGCGAGAAGGAGAAACCCAGCCCTGGAGCCACACATCTGCTCTGAGCAATGGG 50840 
Qy	61	TGCGCCCTCGCAACACACGCCCGCCCTGGCCATGATCAACCCATCAGCGCCCTGCAG 120 
Db	50839	TGCGCCCTCGCAACACACGCCCGCCCTGGCCATGATCAACCCATCAGCGCCCTGCAG 50780 
Qy	121	TCGTCCTTGAACAATCATTGGGCAAGCCAGGAGCCCTTGGCTCAGCTTCTCTCTCTCC 180 
Db	50779	TCGTCCTTGAACAATCATTGGGCAAGCCAGGAGCCCTTGGCTCAGCTTCTCTCTCTCC 50720 
Qy	181	AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAG 240 
Db	50719	AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAG 50660 
Qy	241	CCGGTCTTGAGTCTCGCTCCCAAGAGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGGAG 300 

```
Db 50659 CCGGTCTTGGTCTCTCCACAGGTGTCAGCAGCGGTGTCAGCGGTACTCTGTTGAG 50600
QY 301 AACAGCGATGACCCATGACTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCA 360
Db 50599 AACAGCGATGACCCATGACTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTTCGCA 50540
QY 361 GCACAAATCTTGATGTCCTCCACCTCAGAGCAGCGTCTGTCTGACATCCGCGACATGTC 420
Db 50539 GCACAAATCTTGATGTCCTCCACCTCAGAGCAGCGTCTGTCTGACATCCGCGACATGTC 50480
QY 421 AAAGTCTCTCCCAAGCAGCAGCCCAAGCAGCGTCTCTCTCCAGCGTCTCCCGCCATG 480
Db 50479 AAAGTCTCTCCCAAGCAGCAGCCCAAGCAGCGTCTCTCTCCAGCGTCTCCCGCCATG 50420
QY 481 AAGCTGGAATGATGTCAGCGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGCAT 540
Db 50419 AAGCTGGAATGATGTCAGCGCGCTTTGAGGATGTCCTCAGTGAAGTCTCAACTTTGCAT 50360
QY 541 AAAGAAAGGCGCGCAGTCCAACTGGAATCTCAGCATCTTCTGATCTGAGCCCA 600
Db 50359 AAAGAAAGGCGCGCAGTCCAACTGGAATCTCAGCATCTTCTGATCTGAGCCCA 50300
QY 601 TTTGCTCGAGCTCTCCAGACATCAGAGGCAAAATACCTGCTCTGATCTGAGCCCA 660
Db 50299 TTTGCTCGAGCTCTCCAGACATCAGAGGCAAAATACCTGCTCTGATCTGAGCCCA 50240
QY 661 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720
Db 50239 CAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 50180
QY 721 CTGGCAACGTCAGTACAGCTTAGGAAACGGCGGGGACAAATTTCTGAAAAACATG 780
Db 50179 CTGGCAACGTCAGTACAGCTTAGGAAACGGCGGGGACAAATTTCTGAAAAACATG 50120
QY 781 GACAAAGCCCAACCTCTTTTATGAGTACTGCTGCTCCAGTCTCAGAACCCCTTCT 840
Db 50119 GACAAAGCCCAACCTCTTTTATGAGTACTGCTGCTCCAGTCTCAGAACCCCTTCT 50060
QY 841 ACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGCTTG 900
Db 50059 ACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGCTTG 50000
QY 901 TCAGTGACACGAAAGCAAGGTGAGCAGAGATCTCCCGGTATCTCGGCTCAGAGG 960
Db 49999 TCAGTGACACGAAAGCAAGGTGAGCAGAGATCTCCCGGTATCTCGGCTCAGAGG 49940
QY 961 TCTCAGAAACAATAGCTGCGGAGAGGACAGACTCTAAATCAAGTGAAGTTGTC 1020
Db 49939 TCTCAGAAACAATAGCTGCGGAGAGGACAGACTCTAAATCAAGTGAAGTTGTC 49880
QY 1021 TGTCCGACATTTGTGAGCAAAACATCGCGTAAACCTCCACCTAAAGCAAAACGACAGCAAG 1080
Db 49879 TGTCCGACATTTGTGAGCAAAACATCGCGTAAACCTCCACCTAAAGCAAAACGACAGCAAG 49820
QY 1081 TCACCCGAACCAATTCAGTTTGTAAACAGCGTGAAGATAGCTTCGAGG 1137
Db 49819 TCACCCGAACCAATTCAGTTTGTAAACAGCGTGAAGATAGCTTCGAGG 49763

RESULT 8
AC004505 57493 bp DNA linear PRI 20-JUN-1998
DEFINITION Homo sapiens chromosome 20, p1 clone 86C1 (LBNL H65), complete
sequence.
AC004505.1 GI:2996633
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 57493)
```

```
AUTHORS Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,
Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Miguel,T.,
Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.
Sequencing of human chromosome 20
Unpublished
2 (bases 1 to 57493)
Ricke,D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
3 (bases 1 to 57493)
Gray,J.W., Collins,C., Kimmerly,W., Bondoc,M., Cheng,J.,
Connolly,K.S., Gunning,K.M., Davis,C.A., Kadner,K., Miguel,T.,
Pitluck,S., Pollard,M., Rojeski,H., Subramanian,S. and Martin,C.H.
Direct Submission
Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
4 (bases 1 to 57493)
Kimmerly,W., Bondoc,M., Cheng,J., Connolly,K.S., Gunning,K.M.,
Davis,C.A., Kadner,K., Miguel,T., Pitluck,S., Pollard,M.,
Rojeski,H., Subramanian,S. and Martin,C.H.
Direct Submission
Submitted (30-MAR-1998) Human Genome Center, DOE Joint Genome
Institute, Lawrence Berkeley National Laboratory, MS 74-157,
Berkeley, CA 94720, U.S.A.
5 (bases 1 to 57493)
Ricke,D.O.
Direct Submission
Submitted (20-JUN-1998) DOE Joint Genome Institute
Sequence submitted by:
DOE Joint Genome Institute.
Location/Qualifiers
1. 57493
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q13.2"
/clones="86C1"
/note="LBNL H65"
complement(408..1147)
/rpt_family="LTR1"
repeat_region
complement(1436..1598)
/rpt_family="MIR"
repeat_region
1911..2091
/rpt_family="Alu"
repeat_region
complement(2129..2293)
/rpt_family="L1"
repeat_region
complement(3057..3173)
/note="GRAIL 2 excellent exon, frame 1"
repeat_region
4785..5093
/rpt_family="Alu"
repeat_region
5943..6244
/rpt_family="Alu"
repeat_region
complement(6587..6677)
/note="GRAIL 2 excellent exon, frame 2"
misc_feature
complement(7201..7264)
/note="95% identity B76563"
misc_feature
10223..10384
/note="GRAIL 2 excellent exon, frame 1"
repeat_region
11317..11682
/rpt_family="THE1B"
repeat_region
complement(13299..13486)
/rpt_family="WER3"
repeat_region
complement(13751..14053)
/rpt_family="Alu"
repeat_region
14257..14288
/note="(AC)16"
repeat_region
/rpt_type=tandem
/rpt_unit="ac"
repeat_region
14764..14853
/rpt_family="MIR"
```



```
repeat_region 14896..15223
repeat_region /rpt family="Alu"
               /note="(A)22"
               /rpt_type=tandem
               /rpt_unit="a"
misc_feature complement(15318..15467)
repeat_region 15711..15777
               /rpt family="MIR2"
repeat_region complement(15948..16096)
               /rpt family="MIR"
repeat_region 16477..16516
               /note="(AC)20"
               /rpt_type=tandem
               /rpt_unit="ac"
repeat_region 16785..16810
               /note="(GT)13"
               /rpt_type=tandem
               /rpt_unit="gt"
repeat_region complement(17059..17225)
               /rpt family="MER20"
repeat_region 17821..17854
               /note="(AC)17"
               /rpt_type=tandem
               /rpt_unit="ac"
repeat_region 18459..18498
               /note="(AC)20"
               /rpt_type=tandem
               /rpt_unit="ac"
repeat_region complement(18830..18986)
               /rpt family="MER20"
repeat_region complement(19040..19163)
               /rpt family="MER20"
misc_feature 19538..19743
               /note="GRAIL 2 excellent exon, frame 1"
repeat_region 20903..21214
               /rpt family="Alu"
repeat_region complement(22187..22492)
               /rpt family="Alu"
repeat_region 22935..23051
               /rpt family="Alu"
repeat_region complement(23854..23948)
               /rpt family="MIR"
repeat_region complement(25461..25770)
               /rpt family="Alu"
repeat_region complement(26933..27308)
               /rpt family="THE1"
repeat_region 29135..29159
               /note="(A)25"
               /rpt_type=tandem
               /rpt_unit="a"
repeat_region complement(29380..29538)
               /rpt family="MER20"
misc_feature complement(30961..31186)
               /note="45% protein identity D00570"
STS complement(30967..31290)
               /standard_name="CHIC.GATA45B10.PL5032"
               /note="98% identity"
               /db_xref="dbSTS:G08049"
repeat_region 31020..31079
               /note="(GATA)15"
               /rpt_type=tandem
               /rpt_unit="gata"
repeat_region complement(31986..32185)
               /rpt family="MIR"
repeat_region complement(32446..32626)
               /rpt family="MER20"
misc_feature 32688..32745
               /note="GRAIL 2 excellent exon, frame 2"
STS 35608..35837
               /standard_name="UT238"
               /note="100% identity L31660"

complement(35895..36009)
/rpt family="MERS"
complement(36321..36722)
/rpt family="Alu"
complement(38352..38698)
/rpt family="Alu"
complement(39136..39473)
/rpt family="Alu"
41211..41240
/note="(GT)15"
/rpt_type=tandem
/rpt_unit="gt"
42275..42349
/rpt family="MIR"
43809..44057
/note="GRAIL 2 excellent exon, frame 1"
44378..46447
/note="GRAIL 2 excellent exon, frame 1"
44827..45060
/note="81% identity zv90b08.r1"
/db_xref="dbEST:AA424296"
/standard_name="y170h05"
prim_transcript join(45747..46197,46801..46881)
               /note="98% identity dbEST:H12950, H05537"
               /standard_name="ym34b09.s1"
               /note="97% identity"
46285..46359
/note="82% identity yu60e01.r1"
/db_xref="dbEST:H81145"
46399..46499
/note="83% identity mouse EST MDB1409R"
/db_xref="dbEST:N28081"
46624..46753
/note="GRAIL 2 excellent exon, frame 0"
50132..50197
/note="87% identity af87e05.s1"
/db_xref="dbEST:AA778661"
complement(54507..54580)
/rpt family="MIR"
complement(57082..57397)
/rpt family="Alu"
57082..57103

Query Match 87.3%; Score 1125; DB 9; Length 57493;
Best Local Similarity 99.9%; Pred. No. 2e-306;
Matches 1136; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 GAGGGCAGCGAAGAGAGAAACCCAGGCCCTGGAGCCCAATCTGCTGTGAGCAATGGG 60
DB 45746 GAGGGCAGCGAAGAGAGAAACCCAGGCCCTGGAGCCCAATCTGCTGTGAGCAATGGG 45805
QY 61 TCGGCGCTGCCAACCAACGCGCCCGCCCTGCCATCATCAACCCACTAGCGCCCTGCCAG 120
DB 45806 TCGGCGCTGCCAACCAACGCGCCCGCCCTGCCATCATCAACCCACTAGCGCCCTGCCAG 45865
QY 121 TCGGCTCTCAACAATCACTTGGGCAAGCAGCAGGAGCCCTTGGCTTCACCTTCTCTCTCC 180
DB 45866 TCGGCTCTCAACAATCACTTGGGCAAGCAGCAGGAGCCCTTGGCTTCACCTTCTCTCTCC 45925
QY 181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCCAAGTCCAAAGCAAGAAAGCCAGTCTCCCAA 240
DB 45926 AGCCCAAGTTCAAGCACAATTTCCATGTTCCCAAGTCCAAAGTCCAAAGCAAGAAAGCCAG 45985
QY 241 CCGGCTTTGAGTCTGCTCCCTCCCAAGGTCAGCCAGGCTGCCAGGCGCTACCTGTTTGG 300
DB 45986 CCGGCTTTGAGTCTGCTCCCTCCCAAGGTCAGCCAGGCTGCCAGGCGCTACCTGTTTGG 46045
QY 301 AACAGGATCAGCCCAATTGACCTGACCAAGTCCAAAGCAAGAAAGCCAGTCTCCCAA 360
DB 46046 AACAGGATCAGCCCAATTGACCTGACCAAGTCCAAAGCAAGAAAGCCAGTCTCCCAA 46105
QY 361 GCACAATCTGTATGTCTCCACCTCAGAGCAGCTCTCTGTGACATCCCGACATGGTTC 420
```





Db 2974 TCTCCAGAAACAATAGCTCGGAGAGGACACAGACTCTAAATTCAGGTGTA 3026

|||||

RESULT 10

BC0799550

LOCUS

DEFINITION Mus musculus cDNA clone MGC:90763 IMAGE:6852953, complete cds.

ACCESSION BC0799550

VERSION BC0799550.1 GI:50926038

KEYWORDS MGC.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 4902)

Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Hellon,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.B., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

12477932

2 (bases 1 to 4902)

Director MGC Project.

Direct Submission

Submitted (02-AUG-2004) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)

Tissue Procurement: Dr. James Lin, University of Iowa

cDNA Library Preparation: M. Bento Soares, University of Iowa

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;

Web site: <http://www.nisc.nih.gov/>

Contact: [nisc.mcgenhri.nih.gov](mailto:nisc.mcgenhri.nih.gov)

Akhter,N., Avelle,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsugeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>

Series: IRAK Plate: 176 Row: i Column: 23

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein.

1. .4902

FEATURES

source

/organism="Mus musculus"

/mol\_type="mRNA"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="MGC:90763 IMAGE:6852953"

/tissue\_type="Brain"

/clone\_lib="NIH BMAP\_G10"

/lab\_host="DH10B"

/note="Vector: pYX-ASC"

65..2950

/codon\_start=1

/product="Unknown (protein for MGC:90763)"

/protein\_id="AAH79550.1"

/db\_xref="GI:50926039"

/translation="MPRRKQAPKRAAGYAQAEVLKEBEIEKEEEEEEDSGSAQHO

SSNDGTDEELETGPEQKGYFSCQNSPGSHLSNQDAENESLSDASDOVSDKVCGR

DVSKDANTHPKLPSEPHNCKMDTAVTANILSDYSWGLGGLGFLKLSNERNKCDTLV

ELTVHMTHGYQDDNRKKDKLRPTSYSKPKRAFQDMDKGDAQKVLKCMFCGDSFDS

LQDLVHMIKTKHYQKVPLEFVITSSKMTVPKAKRVFDVNRPCSPDSTTGLADSF

SSQSANQLPNSRYQNGASYTQWQPEACKSQILKCMCGSSHDTLQOLTTHMVT

GHFLAVTSSAKKGQVLDPPLAVEQMOSLSETPNSLSLAPKPSNSRSECTASTTTEL

KKESKKEGEGIEQGVKSEYEDSLQKPLDPTIKYQYLREEDLEDGSKGGDILKS

LENTVTTAINKAQGPSWASPTSHAAATQLSEGTGKPMAMGSQILQIRPNANKURP

IAPKWKMPGLGVPPTSLALYTVQKXETEDDEVKQCKGKESPEHEATSPQEGESFS

KIEPPESRKAPCPLEKBEKPKQKPEPPELVFHKSLHVVDKPVISPTSTRPAASVAH

OSVLNHLKGATEPLRSPSCSPNSSTGPFVFKSLVVDKPVISPTSTRPAASVAH

VLFENTDQPIDLTSKSKRAESSQAQSTSPQKHALCDIADMVKVLPKATPKPAAS

SRVPMKLEIDVRREDVSSEVSTLHKRGQSNWNPQHLLILOQAPASSLPTQSEK

YLLSDLGPOERMQISKFTSLMTTISHLANVKYQLRKTGKTVFLKNDKGRHPIFYCS

DCASQFRTPFTSYISHLESHLGFMQDMTRMAADQSKVQEIISRVSSAQRSPETIAGE

EDTDSKFKCKLCRRFTVSKHAVKHLKSLTHSKSPHHQSQFVADVDEE"

ORIGIN

Query Match 70.6%; Score 909.2; DB 10; Length 4902;

Best Local Similarity 88.4%; Pred. No. 1.6e-245;

Matches 999; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 11 AGAAGAGAGAAACCCAGAGCCCTGGAGCCACACATCTGCTGAGCAATGGGTGCGCCTCG 70

DB 1830 AGAAGAGAGAAACGGAGCCCTTAGAACAGTATCTTCTGACCAATGCTGTGCGCCGG 1889

QY 71 CCAACACAGCCCGCGCTGCGCATGCATCAACCCACTAGCGCCCTGCGAGTCGGTCTGA 130

DB 1890 CTAACACACACCCCGCTCTGCTTCATCAACCCGCTCAGTGCCTGCGAGTCGGTCTGA 1949

QY 131 ACAATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACCTTCTGCTCCAGCCCAAGTT 190

DB 1950 ACAACACCTGGGCAAGCCAGGAGCCCTTGGCTTCCCTTCTGCTCCAGTCCCAACT 2009

QY 191 CAAGCACAAATTTCCATGTTCCACAAAGTCCAAATCTCAATGTGATGGACAAAGCGGTCTTGA 250

DB 2010 CAAGCACAGCCCGCTGTTCCATTAATCAGCCTCCACGTGGTGGACAGCGGTCTATAA 2069

QY 251 GTCCTGCTCCACAAG---GTCCAGCAGCGGTGTCCAGGCGCTACCTGTTTGAGAAACAGCG 307

DB 2070 GTCCACCTCCACAGCGCGCGCGCAGAGACTACTACCTGTTTGAGAAACAGCG 2129

QY 308 ATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTCGCAGCACAAT 367

DB 2130 ACCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTCGCAGCACAAT 2189

QY 368 CTTGTATGTGCCCCACTCTCAGAAGCAGCGTCTGTGTGACATCGCCGACATGGTTCAGAGTCC 427

DB 2190 CTGCACTGTCGCCCACTCTCAGAAGCATGCTCTGTGTGACATGCTGTATATGGTCAAGTCC 2249

QY 428 TCCCAAAAGCCACCCCAAGCCAGCTCTCTCTCCAGGGTCCCCCAATGAAGCTGG 487

DB 2250 TCCCAAAAGCCACCCCAAGCCAGCGCTCTCTCGAGGGTCCCTCTATGAAGCTGG 2309

QY 488 AAATGATGTCAGGCGCTTTGAGGATGTTCCAGTGAAGTCTCAACTTTGTCATAAAGAA 547

DB 2310 AAATGACGTCAGAGCGCTTCGAGGATGTTTCGAGGCAAGTCTCCACTTTGCACAAAAGGA 2369

QY 548 AAGCGCGGAGTCCAACTGGATCCTCAGCATCTTCTGATTTCTACAAGCCCAAGTTGGCT 607  
 Db 2370 AAGCGCGGAGTCCAACTGGATCCTCAGCATCTTCTGATTTCTACAAGCCCAAGTTGGCT 2429

QY 608 CGAGCTCTTCCAGACATCAGAGGCAATACCTGCTGTCTGATCTGGGCCCAAGAGC 667  
 Db 2430 CGAGCTCTTCCAGACATCAGAGGCAATACCTGCTGTCTGATCTGGGCCCAAGAGC 2489

QY 668 GTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGTGGCA 727  
 Db 2490 GGATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGTGGCA 2549

QY 728 AGCTCAAGTACAGCTTAGGAAACGGCGGACAAATTTCTGAAAACATGGCAAG 787  
 Db 2550 AGCTCAATACAGCTTAGGAAACAGGTGGGACAAATTTCTGAAAACATGGCAAG 2609

QY 788 GGCACCCCATCTTTTATTCAGTGAATCTGCTCCAGTTCAGAACCCCTTCTACTACA 847  
 Db 2610 GGCACCCCATCTTTTATTCAGTGAATCTGCTCCAGTTCAGAACCCCTTCTACTACA 2669

QY 848 TCAGTCACTTGAATCTCACTGGGTTTCCAAATGAAGACATGACCCGCTTGTGAGTG 907  
 Db 2670 TCAGTCACTTGAATCTCACTGGGTTTCCAAATGAAGACATGACCCGCTTGTGAGTG 2729

QY 908 ACCACAAGCAAGTGGAGCAAGATCTCCCGGTATCTCGGCTCAGAGGTTCTCCAG 967  
 Db 2730 ACCACAAGCAAGTGGAGCAAGATCTCCCGGTATCTCGGCTCAGAGGTTCTCCAG 2789

QY 968 AAACAATAGTCCGGAAGGACACAGACTCTAAATTCAGTGAATTTGTGCTCGGA 1027  
 Db 2790 AAACAATAGTCCGGAAGGACACAGACTCTAAATTCAGTGAATTTGTGCTCGGA 2849

QY 1028 CATTTGTGAGCAAAATCGGTTAACTCCACTTAAGCAAAACGACAGCAAGTCAACCG 1087  
 Db 2850 CATTTGTGAGCAAAATCGGTTAACTCCACTTAAGCAAAACGACAGCAAGTCAACCG 2909

QY 1088 AACACATTCAGATTTGTAAACAGAGTGGATGAAGATAGTCTGCGAG 1137  
 Db 2910 AACACATTCAGATTTGTAAACAGAGTGGATGAAGATAGTCTGCGAG 2959

RESULT 11  
 AL731822  
 LOCUS Mouse DNA sequence from clone RP23-105H6 on chromosome 2, complete  
 DEFINITION sequence.  
 ACCESSION AL731822  
 VERSION AL731822.31 GI:27368269  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 218037)  
 Hopkins, B.  
 Direct Submission  
 Submitted (19-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Dec 23, 2002 this sequence version replaced gi:2762683.  
 Sequence from the Mouse Genome Sequencing Consortium whole genome  
 shotgun may have been used to confirm this sequence. Sequence data  
 from the whole genome shotgun alone has only been used where it has  
 a phred quality of at least 30.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 -----

During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) RP23-105H6 is from the RP23-23 Mouse PAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

## VECTOR: pBAC3.6.

FEATURES  
source

Location/Qualifiers  
 1. 218037  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="2"  
 /clone="RP23-105H6"  
 /clone\_lib="RPCI-23"

## ORIGIN

Query Match 70.6%; Score 909.2; DB 10; Length 218037;  
 Best Local Similarity 88.4%; Pred. No. 2.6e-245;  
 Matches 999; Conservative 0; Mismatches 128; Indels 3; Gaps 1;

QY 11 AGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGTGGCGCCCTCG 70  
 Db 13029 AGAAGGAGAAACCGAGCCCTTAGACCAATGATCTTCTGACCAATGGTGGCGCGG 13088

QY 71 CCAACCAAGCCCGCCCTGCCATGCAATCAACCACTCAGCGCCCTGAGTCCGTCCTGA 130  
 Db 13089 CTAACCAACCCCGCTCTGCTTCCATCAACCCGCTCAGTGGCCCTGAGTCCGTCCTGA 13148

QY 131 ACATCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTTCTGCTCCAGCCCAAGTT 190  
 Db 13149 ACAACCACTTGGGCAAGCCAGGAGCCCTTGGCTCACTTCTGCTCCAGTCCCACT 13208

QY 191 CAAGCACAATTTCCATGTTCCACAAGTCGAATCTCAATGTCATGACACAAGCGGCTTTGA 250  
 Db 13209 CAAGCAAGCCCGTGTTCATTAATCCAGCTCCAGTGGTGACAAGCGGTCATAA 13268

QY 251 GTCCTGCTCCACAAG---GTCAGCCAGCGTGTCCAGCGCTACCTGTTTGAGAACAGCG 307  
 Db 13269 GTCCACCTCCACAGGCGCGCGCCAGCACTACCTGTTTGAGAACACCG 13328

QY 308 ATGAGCCCATTTGACTGACCAAGTCCAAAGCAAGAGCCGAGTCTCTCGCAGCACAA 367  
 Db 13329 ACCAGCCCATTTGACTGACCAAGTCCAAAGCAAGAGCCGAGTCTCTCGCAGCACAA 13388

QY 368 CTGTGATGTCCCACTCTCAGAACGACGCTCTGTCTGACATCCCGACATGTTCAAGTCC 427  
 Db 13389 CTGTGATGTCCCACTCTCAGAACGACGCTCTGTCTGACATGTTGATATGTTCAAGTCC 13448

QY 428 TCCCAAGCCACCAACCCCAAGCCAGCTCTCTCTCCAGGGTCCCCCCCCATGAAGCTGG 487  
 Db 13449 TCCCAAGCCACCAACCCCAAGCCAGCTCTCTCTCCAGGGTCCCCCCCCATGAAGCTGG 13508

QY 488 AATGAGTGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCACTTTTGCATATAAGAA 547  
 Db 13509 AATGAGTGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCCACTTTTGCATATAAGAA 13568

QY 548 AAGCGCGGAGTCCAACTGGATCCTCAGCATCTTCTGATTTCTACAAGCCCAAGTTGGCT 607  
 Db 13569 AAGCGCGGAGTCCAACTGGATCCTCAGCATCTTCTGATTTCTACAAGCCCAAGTTGGCT 13628





Db 2237 CTAACACACCCCGGCTGCTCCATCAACCCGCTCAGTGCCTCGAGTCGCTCTCA 2296  
Qy 131 ACAATCACTTGGCAAGACGAGCCCTTGCGCTCACCTTCTGCTCCAGCCCAAGTT 190  
Db 2297 ACACCACTTGGCAAGACGAGCCCTTGCGCTCCCTTCTGCTCCAGTCCCAACT 2356  
Qy 191 CAAGCACAATTTCCATGTTCCAAAGTCCAAATCTCAATGTCTATGTCATGGACAAGCCGGTCTTGA 250  
Db 2357 CAAGCAAGCCCGTGTTCATAAATCCAGCTCCACGTGTGGAGACGCGGTATAA 2416  
Qy 251 GTCCTGCTTCCAAAG---GTACGACGAGCGGTGTCAGGCGCTACCTGTTTGGAGAACAGCG 307  
Db 2417 GTCCCACTTCCACAGCGCGCGGCGAGCGGTGGCCAGACACTACCTGTTTGGAGAACAGCG 2476  
Qy 308 ATCAGCGCATTTGACCTGACCAAGTCCAAAGCAAGCAAGCGGAGTCCCTGCGAGCACAAT 367  
Db 2477 ACCAGCGCATTTGACCTGACCAAGTCCAAAGCAAGCAAGCGGAGTCCCTGCGAGCACAAT 2536  
Qy 368 CTTGTATGTCCTCCACCTCAGAACGAGCTCTGTCTGACATCGCGACATGGTCMAAGTCC 427  
Db 2537 CTTGACGCTCCCACTCAGAGCATGCTCTGTGTGACATTTGCTGATATGTTCAAGTCC 2596  
Qy 428 TCCCAAGACCAACCCCAAGCCAGCGCTCTCTTCCAGGGTCCCGCCCATGAAGCTGG 487  
Db 2597 TCCCAAGACCAACCCCAAGCCAGCGCTCTCTGAGGGTCCCTCTATGAAGCTGG 2656  
Qy 488 AATGATGTCAGGCGCTTGGAGTGTCTCCAGTGAAGTCTCAACTTTGTCATTAAGAA 547  
Db 2657 AATGACGCTCAGACGCTTCGAGGATGTTTCGAGCGAAGTCTCCATTTGTCACAAAAGGA 2716  
Qy 548 AAGCGCGCAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTTACAAGCCAGTTTGCCT 607  
Db 2717 AAGCGCGCAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTTACAAGCCAGTTTGCCT 2776  
Qy 608 CGAGCTCTTCCAGACATCAGAGGGCAATACCTGCTGTCTGATCTGGGCCCAAGAGC 667  
Db 2777 CGAGCTCTTCCAGACATCAGAGGGCAATACCTGCTGTCTGATCTGGGCCCAAGAGC 2836  
Qy 668 GTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTACTGGTGGCCA 727  
Db 2837 GGATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCTACTGGTGGCTA 2896  
Qy 728 ACCTCAAGTACCACTTAGGAAAACGGCGGGAACAAAATTTCTGAAAAACATGGACAAAG 787  
Db 2897 ACCTCAATACCACTTAGGAAAACAGGTGGGACAAAGTTCTGAAAAACATGGACAAAG 2956  
Qy 788 GGCACCCCATCTTTTATTCAGTGAAGTGTGCTTCCAGTTCAGAACCCCTTTTACCTACA 847  
Db 2957 GGCACCCCATCTTTTATTCAGTGAAGTGTGCTTCCAGTTCAGAACCCCTTTTACCTACA 3016  
Qy 848 TCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATGACCCCGTGTGAGTGG 907  
Db 3017 TCAGCACCCTTAGAGTCTCACCTGGGTTTCCAAATGAAGACATGACCCCGATGGCGGCTG 3076  
Qy 908 ACCAGCAAGCAGGTGGAGCAGAGATCTCCCGGTTATCGTCCGCTCAGAGGTCTCCAG 967  
Db 3077 ACCAGCAAGCAGGTGGAGCAGAGATCTCCCGAGTGTGCTCGGCTCAGAGGTCTCCAG 3136  
Qy 968 AAACAATAGTCCGGAAGGAGCACAGACTCTAAATTCAAAGTGTAAAGTTGTGCTCGGA 1027  
Db 3137 AAACAATAGTCCGGAAGGAGCACAGACTCTAAATTCAAAGTGTAAAGTTGTGCTCGGA 3196  
Qy 1028 CATTGTGAGCAACATGCGGTAAGTAACTCCACCTTAAGCAAAAGCAGCAAGTCAACCG 1087  
Db 3197 CATTGTGAGCAACATGCGGTAAGTAACTCCACCTTAAGCAAAAGCAGCAAGTCAACCG 3256  
Qy 1088 AACACCAATTACAGTTTGTAAACAGAGCTGGATGAAGAA 1125  
Db 3257 AACACCAATTACAGTTTGTAAACAGAGCTGGATGAAGAA 3294

CQ078477  
LOCUS CQ078477 2012 bp DNA linear PAT 20-JAN-2004  
DEFINITION Sequence 14277 from Patent WO0157278.  
ACCESSION CQ078477  
VERSION CQ078477.1 GI:41049346  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Penn, S.G., Hanzel, D.K., Chen, W. and Rank, D.R.  
TITLE Human genome-derived single exon nucleic acid probes useful for  
analysis of gene expression in human hela cells or other human  
cervical epithelial cells  
JOURNAL Patent: WO 0157278-A 14277 09-AUG-2001;  
FEATURES  
Location/Qualifiers  
1..2012  
/organism="Homo sapiens"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:9606"  
/note="WAP TO AC004505.1-EXPRESSED IN HELA, SIGNAL = 1-NT  
HIT: AF207880.1, EVALUE 0.00e+00-EST\_HUMAN HIT: H12950.1,  
EVALUE 0.00e+00-SWISSPROT HIT: P22265, EVALUE 1.00e-07"  
ORIGIN  
Query Match 51.0%; Score 656.4; DB 6; Length 2012;  
Best Local Similarity 99.8%; Pred. No. 4,7e-174;  
Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GAGGGCAGCGAGAGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60  
Db 1347 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 1406  
Qy 61 TCGCGCCTCGCAACCAACGCGCCCGCTGCGATGCATCAACCCACTAGCGCCCTGCGAG 120  
Db 1407 TCGCGCCTCGCAACCAACGCGCCCGCTGCGATGCATCAACCCACTAGCGCCCTGCGAG 1466  
Qy 121 TCGGTCTTGAACAATCACTTGGGCAAGCAGGAGCCCTTGGCTCAGCTCTCTGCTCC 180  
Db 1467 TCGGTCTTGAACAATCACTTGGGCAAGCAGGAGCCCTTGGCTCAGCTCTCTGCTCC 1526  
Qy 181 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAAGTCCAAATCTCAATGTCATGGACAAG 240  
Db 1527 AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAAGTCCAAATCTCAATGTCATGGACAAG 1586  
Qy 241 CCGGTCTTGAAGTCTGCTCCCAAGAGTTCAGCGGTGTCAGGCGCTTACCTGTTTGAAG 300  
Db 1587 CCGGTCTTGAAGTCTGCTCCCAAGAGTTCAGCGGTGTCAGGCGGTACCTGTTTGAAG 1646  
Qy 301 AACAGGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCAGTCTCTGCGAA 360  
Db 1647 AACAGGATCAGCCCATTTGACCTGACCAAGTCCAAAGCAAGAAAGCCAGTCTCTGCGAA 1706  
Qy 361 GCACAATCTTGATGTGCCCCACTCAGAAGCAGCTCTGCTGACATCCCGACATGGTC 420  
Db 1707 GCACAATCTTGATGTGCCCCACTCAGAAGCAGCTCTGCTGACATCCCGACATGGTC 1766  
Qy 421 AAAGTCTCTCCCCAAAGGCCACCCCAAGCAGCGCTCTCTTCCAGGCTCCCCCCCATG 480  
Db 1767 AAAGTCTCTCCCCAAAGGCCACCCCAAGCAGCGCTCTCTTCCAGGCTCCCCCCCATG 1826  
Qy 481 AAGCTGGAAATGGATGTGAGGCGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTGCAT 540  
Db 1827 AAGCTGGAAATGGATGTGAGGCGCTTTGAGGATGTCTCAGTGAAGTCTCAACTTTGCAT 1886  
Qy 541 AAAAGAAAAGGCGCGCAGTCCCAACTCTCAGCATCTTCTGATTTCTACAAGCCCGAG 600  
Db 1887 AAAAGAAAAGGCGCGCAGTCCCAACTCTCAGCATCTTCTGATTTCTACAAGCCCGAG 1946  
Qy 601 TTTGCTCGAGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTCTCATCTGGGCC 658  
|||||

Db 1947 TTGCTCGAGCCTCTTCCAGACATCAGAGGCAATACCTGCTGTCTGTGATCTGGCCC 2004

RESULT 15  
BV014871 671 bp DNA linear STS 30-MAY-2003  
LOCUS S212P60463FH5.T0 CZECHII/Ei Mus musculus STS genomic, sequence  
DEFINITION tagged site.

ACCESSION BV014871  
VERSION BV014871.1 GI:31098766  
KEYWORDS STS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 671)  
AUTHORS Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K. and Daly, M.J.  
TITLE The mosaic structure of variation in the laboratory mouse genome  
JOURNAL Nature 420 (6915), 574-578 (2002)  
MEDLINE 22354684  
PUBMED 12466852

COMMENT  
Contact: Kerstin Lindblad-Toh  
Whitehead Institute for Biomedical Research, Center for Genome  
Research  
320 Charles Street, Cambridge, MA 02141, USA  
Tel: 6172521477  
Fax: 6172580903  
Email: kersti@genome.wi.mit.edu  
Primer A: None  
Primer B: None  
STS size: 671  
Protocol:  
WGS-discovery: Paired-end low-coverage whole genome shotgun reads  
were generated from 12981/SvImJ, C3H/HeJ, and BALB/cByJ. The WGS  
reads were placed uniquely on the MGSv3 C57BL/6J assembly and SNP  
detection was carried out by SSAHA-SNP. 225,000 reads were  
annotated  
as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J  
and the strain from which the particular read came. The validation  
rate for these SNPs was estimated at approximately 98%.

FEATURES  
source  
1..671  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="CZECHII/Ei"  
/db\_xref="taxon:10090"  
/map="4 2 22-621 170891742-170892341"  
/clone\_lib="CZECHII/Ei"  
<1..>671

STS  
ORIGIN

Query Match 44.1%; Score 568.6; DB 11; Length 671;  
Best Local Similarity 90.5%; Pred. No. 2.8e-149;  
Matches 607; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 363 ACAATCTTGTATGTCCTCCACCTCAGAAGCAGCCTCTGTGACATCGCGGACATGGTCAA 422  
Db 1 ACAATCTGACGTCCTCCACCTCAGAAGCAGCCTCTGTGACATGCGGATGGTCAA 60

Qy 423 AGTCTCTCCCAAGCACCACCCCAAGCCAGCCTCTCTCCAGGGTCCCTCCCATGAA 482  
Db 61 GGTCTCTCCCAAGCACCACCCCAAGCCAGCCTCTCTCCAGGGTCCCTCTCATGAA 120

Qy 483 GCTGGAATGGATGTGAGCGGCTTGGAGATGTCTCCAGTGAAGTCTCAACTTTGCATAA 542  
Db 121 GCTGGAATTCGACGTCGACGCTTTGAGGATGTTTCGAGCGAAGTCTCCACTTGCACAA 180

Qy 543 AAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTACAAGCCAGTT 602  
Db 181 AAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTTACAAGCCAGTT 240

Qy 603 TGCTCGAGCCTCTTCCAGACATCAGAGGCAATACCTGCTGTCTGTGATCTGGCCCCACA 662  
Db 241 TGCTCGAGCCTCTTCCAGACCTCAGAGGGGCAATACCTGCTTTTCTGACCTGGGCCCCCA 300

Qy 663 AGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGCT 722  
Db 301 AGAGCGGATGCAATCTCCAAGTTTACCGGACTCTCCATGACCAATCAGCCACTGGCT 360

Qy 723 GGCCAACGTCAAGTACCAAGTTTAGGAAAAAGGGGGGACAAAAATTTCTGAAAAACATGGA 782  
Db 361 GGCTTAACGTCAAAATACCAAGTTTAGGAAAAAGGGTGGGACAAAGTTCTCTGAAAAACATGGA 420

Qy 783 CAAGGGCCACCCCATCTTTTATTCAGTGTGCTCCAGTTCAGAACCCCTTCTAC 842  
Db 421 CAAGGGGACCCCATCTTTTACTGAGTGTGCTCCAGTTCAGAACCCCTTCTAC 480

Qy 843 CTACATCAGTCACCTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATGACCCGCTTGC 902  
Db 481 CTACATCAGCCACCTTAGAGTCTCACCTGGGTTTCCAAATGAAGGACATGACCCGATGGC 540

Qy 903 AGTGGACCAAGCAAGCAAGTGGAGCAAGAGATCTCCCGGATTCGTCGGTCAAGAGTC 962  
Db 541 GGCTGACCAAGCAAGTGGAGCAAGAGATCTCCCGAGTGTCTCGGCTCAGAGTC 600

Qy 963 TCCAGAAACAATAGCTGCCGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTGCTG 1022  
Db 601 TCCAGAAACAATAGCTGCCGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTGCTG 660

Qy 1023 TCGGACATTTG 1033  
Db 661 TCGGACATTTG 671

Search completed: May 5, 2005, 03:58:23  
Job time : 6299.78 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:02:41 ; Search time 891.572 Seconds  
(without alignments)  
8551.893 Million cell updates/sec

Title: US-08-731-499-5  
Perfect score: 1288  
Sequence: 1 GAGGCGACGAGAGGAGAA.....AAAAAATTC 1288

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04:  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	100.0	1288	2	AAV04700 Homo sapi
2	1268	98.4	3313	10	Abz79883 Human nuc
3	1265	98.2	3322	10	ADB62219 Human CDN
4	1013.6	78.7	3442	10	ADC30023 Human nov
5	773	60.0	2449	4	AAS26449 Human cDN
6	773	60.0	2449	8	Abx73790 Human nov
7	770.4	59.8	799	4	AAS26014 Human CDN
8	770.4	59.8	799	8	Abx73355 Human nov
9	656.4	51.0	2012	4	AAI24344 Probe #14
10	521	40.5	1964	4	AAI15164 Probe #50
11	390.4	30.3	3604	4	AAI2638 Human bre
12	390.4	30.3	3604	4	AAI06732 Human rep
13	390.4	30.3	3907	6	ABK34640 Human CDN
14	390.4	30.3	7081	10	ADF55427 Human DNA
15	390	30.3	1079	10	ADC31952 Human nov
16	387.2	29.3	4854	13	ADRO6933 Full leng
17	377.8	29.3	2918	2	AXA40074 Colon can
18	272	21.1	3692	12	ADQ20133 Human sof
19	272	21.1	4407	5	ABX71366 Human tra
20	272	21.1	5065	12	ADQ23080 Human sof

21	272	21.1	5065	12	ADQ24468 Human sof
22	272	21.1	5065	12	ADQ24179 Human sof
23	256.4	19.9	1144	2	AAV61044 Nucleotid
c 24	157	12.2	346	6	ABK64270 Human ben
25	143.8	11.2	1999	4	AAK85937 Human imm
26	143.4	11.1	2044	4	AAS25958 Human cDN
27	143.4	11.1	2044	8	ABX73299 Human nov
28	143	11.1	852	4	AAS26397 Human CDN
29	143	11.1	852	8	ABX73738 Human nov
30	105	8.2	3370	10	ADB62184 Human CDN
31	97.2	7.5	450	9	ACH12952 Human adu
32	65.2	5.1	537	9	ACH32461 Human end
33	63.8	5.0	1258	4	AAK85936 Human imm
34	63.2	4.9	555	4	AAK63792 Human imm
35	51.6	4.0	636	4	AAI15974 Human bre
36	51.6	4.0	640	11	ACN92451 Breast ca
37	51.6	4.0	651	4	AAI24818 Human bre
38	51.6	4.0	904	11	ACN85935 Breast ca
c 39	46.4	3.6	2000	8	ADA71938 Rice gene
40	41.6	3.2	258	11	ADI31447 Human CDN
c 41	40	3.1	515	12	ACH73311 Human gen
42	40	3.1	2000	8	ADA71938 Rice gene
43	38.6	3.0	6643	12	ADQ83730 Human tum
44	38.2	3.0	677	2	AAK39783 Gastric c
45	38.2	3.0	4590	5	AAH24065 Yeast AOD

ALIGNMENTS

RESULT 1  
AAV04700  
ID AAV04700 standard; cDNA; 1288 BP.  
XX  
AC AAV04700;  
XX  
DT 21-JUL-1998 (first entry)  
XX  
DE Homo sapiens 20q13 amplicon 41.1 transcript.  
XX  
KW 20q13 amplicon; chromosome 20; tumour; detection; homeobox T shirt;  
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;  
KW treatment; age-related macular degeneration; retinitis pigmentation;  
KW Leber's congenital amaurosis; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO9802539-A1.  
XX  
PD 22-JAN-1998.  
XX  
PF 15-JUL-1997; 97WO-US012343.  
XX  
PR 15-JUL-1996; 96US-00680395.  
PR 16-OCT-1996; 96US-00731499.  
PR 17-JAN-1997; 97US-00785532.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;  
XX WPI; 1998-110587/10.  
XX  
DR New sequences from the 20q13 amplicon - used for detecting chromosomal  
PT abnormalities, particularly tumours, and for developing products for  
PT treating diseases.  
XX  
PS Claim 1; Page 62; 91pp; English.  
XX  
CC The sequence is that of a cDNA sequence 41.1, which was isolated from the  
CC 20q13 amplicon. It shows homology to the homeobox T shirt gene in  
CC Drosophila. It can be used as a probe for the detection of chromosomal  
CC abnormalities at 20q13. It and other sequences isolated from the 20q13

CC	amplicon are consistently amplified in primary tumours. These sequences	
CC	are useful as probes or as probe targets for monitoring the relative copy	
CC	number of corresponding sequences from a biological sample such as tumour	
CC	cells. The sequences can also be used in therapeutic applications for	
CC	modulating the expression of the endogenous gene or the activity of the	
CC	gene product. Examples of therapeutic approaches include antisense	
CC	inhibition of gene expression, gene therapy, and monoclonal antibodies	
CC	that specifically bind the gene products. The products can also be used	
CC	in the treatment of other diseases, e.g. age-related macular	
CC	degeneration, Leber's congenital anaurosis and retinitis pigmentation	
XX		
SQ	Sequence 1288 BP; 355 A; 396 C; 274 G; 263 T; 0 U; 0 Other;	
	Query Match 100.0%; Score 1288; DB 2; Length 1288;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 1288; Conservative 0; Mismatches. 0; Indels 0; Gaps 0;	
QY	1 GAGGGCAGCGAGAGAGAAACCCCGCCCTCGGAGCCCAATCTCTGAGCAATGGG 60	
DB	1 GAGGGCAGCGAGAGAGAAACCCCGCCCTCGGAGCCCAATCTCTGAGCAATGGG 60	
QY	61 TGGCCCTCGCCCAACCAACCGCCCGCCCTGCCATGTCATCAACCACTCAGCGCCCTGCAG 120	
DB	61 TGGCCCTCGCCCAACCAACCGCCCGCCCTGCCATGTCATCAACCACTCAGCGCCCTGCAG 120	
QY	121 TCCGTCCTGAAACAATCACTTTGGGCAAGGCCAGGAGCCCTTGGGCTCACCTTCTGCTCC 180	
DB	121 TCCGTCCTGAAACAATCACTTTGGGCAAGGCCAGGAGCCCTTGGGCTCACCTTCTGCTCC 180	
QY	181 AGCCCAAGTTCAAGCAACAATTTCCATGTTCCACAAGTCGAATCTCAATGTGATGGACAAG 240	
DB	181 AGCCCAAGTTCAAGCAACAATTTCCATGTTCCACAAGTCGAATCTCAATGTGATGGACAAG 240	
QY	241 CCGGTCCTGAGTCCTGGCTCCACAAGTCCAGGTCAGCGGTGTCAGGCGCTACCTGTTGAG 300	
DB	241 CCGGTCCTGAGTCCTGGCTCCACAAGTCCAGGTCAGCGGTGTCAGGCGCTACCTGTTGAG 300	
QY	301 AACAGCGATCAGCGCCATTTGACCTGACCAAGTCCAAAAGCAAGAAAGCGAGTCTCGCAA 360	
DB	301 AACAGCGATCAGCGCCATTTGACCTGACCAAGTCCAAAAGCAAGAAAGCGAGTCTCGCAA 360	
QY	361 GCACAAATCTTGATGTCCTCCCACTCAGAACGACGCTCTGTGTCATGTCGCGACATGGTC 420	
DB	361 GCACAAATCTTGATGTCCTCCCACTCAGAACGACGCTCTGTGTCATGTCGCGACATGGTC 420	
QY	421 AAGTCTCTCCCAAGCCACCAACCGCCCTCTCTCTCAGGTCCTCCCGCCCATG 480	
DB	421 AAGTCTCTCCCAAGCCACCAACCGCCCTCTCTCTCAGGTCCTCCCGCCCATG 480	
QY	481 AAGCTGGAATGGATGTCCAGCGCCCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAAT 540	
DB	481 AAGCTGGAATGGATGTCCAGCGCCCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAAT 540	
QY	541 AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTGATCTTACAAGCCCAAG 600	
DB	541 AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTGATCTTACAAGCCCAAG 600	
QY	601 TTTGCTCTCGAGCCTCTTCCAGATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660	
DB	601 TTTGCTCTCGAGCCTCTTCCAGATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCA 660	
QY	661 CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720	
DB	661 CAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720	
QY	721 CTGGCCAACTCAAGTACCAAGTCTAGAAAACCGGCGGGGCAAAAATTTCTGAAAAACATG 780	
DB	721 CTGGCCAACTCAAGTACCAAGTCTAGAAAACCGGCGGGGCAAAAATTTCTGAAAAACATG 780	
QY	781 GACAAAGGCCACCCCATCTTTTATTCAGTGAATCTGTCCTCCAGTTCAGAACCCCTTCT 840	
DB	781 GACAAAGGCCACCCCATCTTTTATTCAGTGAATCTGTCCTCCAGTTCAGAACCCCTTCT 840	

RESULT 2

ABZ79883  
ID ABZ79883 standard; cDNA; 3313 BP.  
XX

AC ABZ79883;  
XX

DT 19-MAY-2003 (first entry)  
XX

DE Human nucleic-acid associated protein 11 cDNA SEQ ID NO:44.  
XX

KW Human; nucleic-acid associated protein; NAAP; cardiant; cytostatic;  
KW neuroprotective; gene therapy; cardiovascular disorder; cancer;  
KW neurological disorder; gene, ss.  
XX

OS Homo sapiens.  
XX

FH Key Location/Qualifiers  
FT CDS 69..3173  
FT /tag= a  
FT /product= "nucleic-acid associated protein 11"

XX WO2003016549-A2.  
XX

PN 27-FEB-2003.  
XX

PD 14-AUG-2002; 2002WO-US025829.  
XX

XX 17-AUG-2001; 2001US-031311P.  
XX

PR 24-AUG-2001; 2001US-0314682P.  
XX

PR 24-AUG-2001; 2001US-0314756P.  
XX

PR 27-AUG-2001; 2001US-0315105P.  
XX

PR 31-AUG-2001; 2001US-0316751P.  
XX

PR 31-AUG-2001; 2001US-0316856P.  
XX

PR 05-OCT-2001; 2001US-0328185P.  
XX

PA (INCY-) INCYTE GENOMICS INC.  
XX

XX Barroso I, Baughn MR, Becha SD, Blake JJ, Borowsky ML, Burford N;  
PI Duggan BM, Elliott VS, Emerling BM, Forsythe IJ, Gietzen KJ;  
PI Gorvad AE, Griffin JA, Hafalia AJA, Honchell CD, Ison CH, Khan PA;

	Query Match	98.4%;	Score 1268;	DB 10;	Length 3313;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 1268;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	GAGGGCAGCAGAGGAGAAACCCAGCCCTCGAGCGCCACATCTGCTCTGAGCAATGGG	60		
Db	2046	GAGGGCAGCAGAGGAGAAACCCAGCCCTCGAGCGCCACATCTGCTCTGAGCAATGGG	2105		
Qy	61	TGCGCCCTCGCCAAACACAGCGCCCGGCCCTGCGCATCAACCCACATCAGCGCCCTCGAG	120		
Db	2106	TGCGCCCTCGCCAAACACAGCGCCCGGCCCTGCGCATCAACCCACATCAGCGCCCTCGAG	2165		
Qy	121	TCGTCCTCGAAATCATCTTGGGCAAGCAGGAGCCCTTGCGCTCACCTTCTGTCTCC	180		
Db	2166	TCCGTCCTGAACATCATCTTGGGCAAGCAGGAGCCCTTGCGCTCACCTTCTGTCTCC	2225		
Qy	181	AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAAAGTCGAATCTCAATGTCATGCACAAG	240		
Db	2226	AGCCCAAGTTCAAGCACAATTTCCATGTTCCAAAGTCGAATCTCAATGTCATGCACAAG	2285		
Qy	241	CCGGTCTTGAGTCTCGCTCCACAAAGGTTCAGCAGCGGTGTCCAGCGCGTACTCTGTTGAG	300		
Db	2286	CCGGTCTTGAGTCTCGCTCCACAAAGGTTCAGCAGCGGTGTCCAGCGCGTACTCTGTTGAG	2345		
Qy	301	AACAGCGATCAGCCCATTTGACCTGACCAAGTTCAAAAGCAAGAAGCGGAGTCTTCGAA	360		
Db	2346	AACAGCGATCAGCCCATTTGACCTGACCAAGTTCAAAAGCAAGAAGCGGAGTCTTCGAA	2405		
Qy	361	GCACAAATCTGTATGTCCTCCACCTTCAGAGCAGCGCTGTCTGATCATCGCGCATGGTC	420		
Db	2406	GCACAAATCTGTATGTCCTCCACCTTCAGAGCAGCGCTGTCTGATCATCGCGCATGGTC	2455		
Qy	421	AAAGTCTCTCCCAAGGCACACACCCCAAGCAGCGCTCTCTCCAGGGTCCCCCCCATG	480		
Db	2466	AAAGTCTCTCCCAAGGCACACACCCCAAGCAGCGCTCTCTCCAGGGTCCCCCCCATG	2525		
Qy	481	AAGCTGGAATPGATGTGAGCGGTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT	540		
Db	2526	AAGCTGGAATPGATGTGAGCGGTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT	2585		
Qy	541	AAAAGAAAAGCCGGCAGTCCAACTGGNATCCTCAGCATCTTCTGATCTCAGAGCCCGAG	600		
Db	2586	AAAAGAAAAGCCGGCAGTCCAACTGGNATCCTCAGCATCTTCTGATCTCAGAGCCCGAG	2645		
Qy	601	TTTGCTCTGAGCCCTCTTCCAGACATCAGAGGGCAAAATACCTTGCTGTCTGATCTGGGCCA	660		
Db	2646	TTTGCTCTGAGCCCTCTTCCAGACATCAGAGGGCAAAATACCTTGCTGTCTGATCTGGGCCA	2705		
Qy	661	CAAGAGCGTATGCAAAATCTCTAAGTTTTCAGGGAATCTCAATGACCACTATCAGTCACTGG	720		





RESULT 5  
AAS26449  
ID AAS26449 standard; cDNA; 2449 BP.  
XX  
AC AAS26449;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human cDNA encoding a novel secreted protein, Seq ID 628.  
XX  
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmologic; vulnary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO20015322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001341.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225219P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225466P.  
PR 14-AUG-2000; 2000US-0225467P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.

PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249246P.  
 PR 17-NOV-2000; 2000US-0249247P.  
 PR 17-NOV-2000; 2000US-0249248P.  
 PR 17-NOV-2000; 2000US-0249249P.  
 PR 17-NOV-2000; 2000US-0249250P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251989P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251472P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-488783/53.  
 DR P-PSDB; AAU16462.  
 XX  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives.  
 XX  
 PS Claim 1; SEQ ID NO 628; 980pp; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 CC in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Antibodies to the proteins can also be used in  
 CC alleviating symptoms associated with the disorders and in diagnostic  
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
 CC and many other disorders listed in the specification. The polypeptides  
 CC can also be used to aid wound healing and epithelial cell proliferation,  
 CC to prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence encodes a novel secreted protein of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 Query Match 60.0%; Score 773; DB 4; Length 2449;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-216;  
 Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 514 GTCTCCAGTGAAGTCTCAACTTTGGATTAAGAAAGGCGGCGAGTCCCACTGGAATCCT 573  
 DB 95 GTCTCCAGTGAAGTCTCAACTTTGGATTAAGAAAGGCGGCGAGTCCCACTGGAATCCT 154  
 QY 574 CAGCATCTTCTGATCTCAAGCCAGTTTGGCTCGAGCCTTCCAGACATCAGAGGGC 633  
 DB 155 CAGCATCTTCTGATCTCAAGCCAGTTTGGCTCGAGCCTTCCAGACATCAGAGGGC 214  
 QY 634 AAATACCTGCTGCTGATCTGGGCCCAAGAGCGTATGCAAAATCTTAAGTTTACGGGA 693

Db 215 AATACCTGCTGCTGATCTGGGCCCAAGAGCGTATGCAAAATCTCTAAGTTTACGGGA 274  
 QY 694 CTCTCAATGACCACCTATCAGTCACTGGCTGGCCCAAGTCAAGTACCAGCTTAGGAAAAACG 753  
 Db 275 CTCTCAATGACCACCTATCAGTCACTGGCTGGCCCAAGTCAAGTACCAGCTTAGGAAAAACG 334  
 QY 754 GCGGGACAAAATTTCTGAAAAAATGAGCAAAAGCCCAACCCCATCTTTTATTCAGTGCAC 813  
 Db 335 GCGGGACAAAATTTCTGAAAAAATGAGCAAAAGCCCAACCCCATCTTTTATTCAGTGCAC 394  
 QY 814 TGTGCTCCCAAGTTTCTGAAAAAATGAGCAAAAGCCCAACCCCATCTTTTATTCAGTGCAC 873  
 Db 395 TGTGCTCCCAAGTTTCTGAAAAAATGAGCAAAAGCCCAACCCCATCTTTTATTCAGTGCAC 454  
 QY 874 TTCAAATGAAGAGCATGACCCGCTTGTCACTGAGCAAGCAAGAGGAGCAAGAG 933  
 Db 455 TTCAAATGAAGAGCATGACCCGCTTGTCACTGAGCAAGCAAGAGGAGCAAGAG 514  
 QY 934 ATCTCCCGGGTATCGTGGCTCAGAGGTTCTCCAGAAACAATAGCTGCCGAGAGGACACA 993  
 Db 515 ATCTCCCGGGTATCGTGGCTCAGAGGTTCTCCAGAAACAATAGCTGCCGAGAGGACACA 574  
 QY 994 GACTCTAAATTCAGTGTAAAGTGTGCTGCGGCACTTTGTGAGCAAAACATGCGGTAAAA 1053  
 Db 575 GACTCTAAATTCAGTGTAAAGTGTGCTGCGGCACTTTGTGAGCAAAACATGCGGTAAAA 634  
 QY 1054 CTCACCTTAAGCAAAAGCAGCAAGTCACTCCGCAACCACTTTCACAGTTTGTAAACAGAC 1113  
 Db 635 CTCACCTTAAGCAAAAGCAGCAAGTCACTCCGCAACCACTTTCACAGTTTGTAAACAGAC 694  
 QY 1114 GTGATGAAGAAATAGCTCTGAGGACGAAATGCTTAGTTTTCAGCTTTCAGCTGGATCC 1173  
 Db 695 GTGATGAAGAAATAGCTCTGAGGACGAAATGCTTAGTTTTCAGCTTTCAGCTGGATCC 754  
 QY 1174 CCTCAGCTGAACCTTCTTCTGCGGCACTTCTGAGCACTTTCAGCACTTGAATGAAC 1233  
 Db 755 CCTCAGCTGAACCTTCTTCTGCGGCACTTCTGAGCACTTTCAGCACTTGAATGAAC 814  
 QY 1234 CCTCCTGACACCTTGGCTCTGAGAGAGTGTGCAAAAGGAGGAGGAGGAGGAGGAGGAG 1286  
 Db 815 CCTCCTGACACCTTGGCTCTGAGAGAGTGTGCAAAAGGAGGAGGAGGAGGAGGAGGAG 867  
 RESULT 6  
 ABX73790  
 ID ABX73790 standard; DNA; 2449 BP.  
 XX  
 AC ABX73790;  
 XX  
 DT 18-MAR-2003 (first entry)  
 XX  
 DE Human novel polynucleotide #618.  
 XX  
 KW Human; gene; ds; neural disorder; immune system disorder; renal disorder;  
 KW muscular disorder; respiratory disease; reproductive disorder;  
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;  
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
 KW haemostatic; antiarteriosclerotic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002132753-A1.  
 XX  
 PD 19-SEP-2002.  
 XX  
 PF 17-JAN-2001; 2001US-00764864.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 07-JUL-2000; 2000US-0216647P.

PR	07-JUL-2000;	2000US-0216880P.	CC	leukaemia), inflammatory diseases (e.g. septic shock, buritis and
PR	11-JUL-2000;	2000US-0217487P.	CC	appendicitis), allergic reactions and conditions (e.g. asthma), blood
PR	11-JUL-2000;	2000US-0217496P.	CC	related disorders (e.g. thrombosis, atherosclerosis and myocardial
PR	14-JUL-2000;	2000US-0218290P.	CC	infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
PR	26-JUL-2000;	2000US-0220963P.	CC	human novel polynucleotides of the invention
PR	26-JUL-2000;	2000US-0220964P.	XX	
PR	14-AUG-2000;	2000US-0224518P.	SQ	Sequence 2449 BP; 746 A; 526 C; 477 G; 687 T; 0 U; 13 Other;
PR	14-AUG-2000;	2000US-0224519P.		
PR	14-AUG-2000;	2000US-0225267P.		
PR	14-AUG-2000;	2000US-0225268P.		
PR	14-AUG-2000;	2000US-0225270P.		
PR	14-AUG-2000;	2000US-0225447P.		
PR	14-AUG-2000;	2000US-0225757P.		
PR	14-AUG-2000;	2000US-0225758P.		
PR	22-AUG-2000;	2000US-0226868P.		
PR	30-AUG-2000;	2000US-0228924P.		
PR	01-SEP-2000;	2000US-0229287P.		
PR	01-SEP-2000;	2000US-0229343P.		
PR	01-SEP-2000;	2000US-0229344P.		
PR	01-SEP-2000;	2000US-0229345P.		
PR	05-SEP-2000;	2000US-0229509P.		
PR	05-SEP-2000;	2000US-0229513P.		
PR	08-SEP-2000;	2000US-0231413P.		
PR	21-SEP-2000;	2000US-0234223P.		
PR	21-SEP-2000;	2000US-0234274P.		
PR	25-SEP-2000;	2000US-0234997P.		
PR	27-SEP-2000;	2000US-0235834P.		
PR	29-SEP-2000;	2000US-0236327P.		
PR	29-SEP-2000;	2000US-0236367P.		
PR	29-SEP-2000;	2000US-0236368P.		
PR	29-SEP-2000;	2000US-0236369P.		
PR	29-SEP-2000;	2000US-0236370P.		
PR	02-OCT-2000;	2000US-0236802P.		
PR	02-OCT-2000;	2000US-0237037P.		
PR	02-OCT-2000;	2000US-0237038P.		
PR	02-OCT-2000;	2000US-0237039P.		
PR	02-OCT-2000;	2000US-0237040P.		
PR	13-OCT-2000;	2000US-0239315P.		
PR	20-OCT-2000;	2000US-0240960P.		
PR	20-OCT-2000;	2000US-0241785P.		
PR	20-OCT-2000;	2000US-0241809P.		
PR	01-NOV-2000;	2000US-0244617P.		
PR	17-NOV-2000;	2000US-0249299P.		
PR	08-DEC-2000;	2000US-0251856P.		
PR	08-DEC-2000;	2000US-0251868P.		
PR	08-DEC-2000;	2000US-0251869P.		
XX				
PA	(ROSE/) ROSEN C A.			
PA	(RUBE/) RUBEN S M.			
PA	(BARA/) BARASH S C.			
PI	Rosen CA, Ruben SM, Barash SC;			
XX				
DR	WPI; 2003-147444/14.			
DR	P-PSDB; ABUS5530.			
XX				
PT	New polypeptides and nucleic acids, useful in gene therapy for treating,			
PT	inhibiting or preventing e.g. neural, immune system, muscular,			
PT	respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or			
PT	renal disorders.			
XX				
PS	Claim 1; SEQ ID NO 628; 402pp; English.			
XX				
CC	The invention relates to human novel polypeptides and their associated			
CC	polynucleotides. The polypeptides and polynucleotides are useful in gene			
CC	therapy for treating, inhibiting or preventing neural disorders, immune			
CC	system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis			
CC	and multiple sclerosis), muscular disorders, respiratory diseases (e.g.			
CC	nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,			
CC	gastrointestinal disorders, pulmonary disorders, cardiovascular disorders,			
CC	(e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left			
CC	heart syndrome), renal disorders (e.g. acute kidney failure and end-stage			
CC	renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and			

CC	leukaemia), inflammatory diseases (e.g. septic shock, buritis and
CC	appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC	related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC	infarction) and cancerous diseases. Sequences ABX73173-ABX74167 represent
CC	human novel polynucleotides of the invention
XX	
SQ	Sequence 2449 BP; 746 A; 526 C; 477 G; 687 T; 0 U; 13 Other;
	Query Match
	Best Local Similarity 60.0%; Score 773; DB 8; Length 2449;
	Matches 773; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	514 GTCTCAGTGAAGTCTCAACTTTTGATATAAAGGCGCGCAGTCCAACTGGATCCT 573
DB	95 GTCTCAGTGAAGTCTCAACTTTTGATATAAAGGCGCGCAGTCCAACTGGATCCT 154
QY	574 CAGCATCTTCTGATTTCTACAAGCCAGTTTGCTCGAGCTCTTCAGACATCAGAGGCG 633
DB	155 CAGCATCTTCTGATTTCTACAAGCCAGTTTGCTCGAGCTCTTCAGACATCAGAGGCG 214
QY	634 AATACCTGCTGCTGATCTGGGCCCACAAGAGCGTATGCAAAATCTTAAGTTTACGGGA 693
DB	215 AATACCTGCTGCTGATCTGGGCCCACAAGAGCGTATGCAAAATCTTAAGTTTACGGGA 274
QY	694 CTCTCAATGACCACATCAGTCACTGGCTGGCCAACTCAAGTACCAGCTTAGGAAAACG 753
DB	275 CTCTCAATGACCACATCAGTCACTGGCTGGCCAACTCAAGTACCAGCTTAGGAAAACG 334
QY	754 GCGGGACAAAATTTCTGAAAAACATGGACAAGGCGCCATCTTTTATTGCACTGAC 813
DB	335 GCGGGACAAAATTTCTGAAAAACATGGACAAGGCGCCATCTTTTATTGCACTGAC 394
QY	814 TGTGCTCCAGTTCAGAACCCCTTCTACCTACATCAGTCACTTGAATCTCACCTGGGT 873
DB	395 TGTGCTCCAGTTCAGAACCCCTTCTACCTACATCAGTCACTTGAATCTCACCTGGGT 454
QY	874 TTCCAAATGAAGGACATGACCCGCTTGTGAGTGGACCAAGCAAGAGTGGAGCAAGAG 933
DB	455 TTCCAAATGAAGGACATGACCCGCTTGTGAGTGGACCAAGCAAGAGTGGAGCAAGAG 514
QY	934 ATCTCCCGGGTATCGTGGCTCAGAGGTTCTCAGAAAAAATAGTGTGCCGAGAGGACACA 993
DB	515 ATCTCCCGGGTATCGTGGCTCAGAGGTTCTCAGAAAAAATAGTGTGCCGAGAGGACACA 574
QY	994 GACTCTAAATTCAGTGTAAAGTTGTGCTGCGACATTTGTGAGCAAAACATGCGGTAAAA 1053
DB	575 GACTCTAAATTCAGTGTAAAGTTGTGCTGCGACATTTGTGAGCAAAACATGCGGTAAAA 634
QY	1054 CTCCACCTTAAGCAAAACGACAGCAAGTCAACCGCAACACCATTTCAAGTTTGTAAACAGAC 1113
DB	635 CTCCACCTTAAGCAAAACGACAGCAAGTCAACCGCAACACCATTTCAAGTTTGTAAACAGAC 694
QY	1114 GTGGATGAAGAATAGCTCTGCGAGGACGAATGCTTAGTTTCCACTTTCCAGCTTGATCC 1173
DB	695 GTGGATGAAGAATAGCTCTGCGAGGACGAATGCTTAGTTTCCACTTTCCAGCTTGATCC 754
QY	1174 CTTCACTACGACCCCTTCTGCTGCGACCATCTGCTCTGACATTTGAATTTCACTTCAACT 1233
DB	755 CTTCACTACGACCCCTTCTGCTGCGACCATCTGCTCTGACATTTGAATTTCACTTCAACT 814
QY	1234 CTTCTGTGACACCCCTGCTCTGAGAGACATGCGCAAAAAAATAAAAAAATAAT 1286
DB	815 CTTCTGTGACACCCCTGCTCTGAGAGACATGCGCAAAAAAATAAAAAAATAAT 867
	RESULT 7
ID	AAS26014
XX	AAS26014 standard; cDNA; 799 BP.
AC	AAS26014;
XX	
DT	07-NOV-2001 (first entry)
XX	



DE Human cDNA encoding a novel secreted protein, Seq ID 193.  
XX  
KW Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic;  
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;  
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;  
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;  
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
KW cerebral ischaemia; angiogenesis; nervous system disorder;  
KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
KW wound healing; epithelial cell proliferation; skin ageing; food additive;  
KW preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200153322-A2.  
XX  
XX  
PD 02-AUG-2001.  
XX  
XX  
PF 17-JAN-2001; 2001WO-US001341.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0218647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 06-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0233397P.  
PR 14-SEP-2000; 2000US-0233398P.  
PR 14-SEP-2000; 2000US-0233399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249254P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249257P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.

05-DEC-2000; 2000US-0251988P.  
06-DEC-2000; 2000US-0256719P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-488783/53.  
DR P-PSDB; AAU16027.  
New nucleic acid molecules encoding 461 human secreted proteins for  
diagnosing, preventing, treating or ameliorating medical conditions and  
used as food additives or preservatives.  
Claim 1; SEQ ID NO 193; 980pp; English.  
The invention relates to isolated nucleic acid molecules and their  
encoded secreted proteins. The nucleic acids and proteins are used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
in diagnosing a pathological condition or susceptibility to a  
pathological condition. Antibodies to the proteins can also be used in  
alleviating symptoms associated with the disorders and in diagnostic  
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays  
(ELISA). Disorders which are diagnosed or treated include autoimmune  
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,  
nervous system disorders e.g. Alzheimer's disease, infections caused by  
bacteria, viruses and fungi and ocular disorders e.g. corneal infection,  
and many other disorders listed in the specification. The polypeptides  
can also be used to aid wound healing and epithelial cell proliferation,  
to prevent skin aging due to sunburn, to maintain organs before  
transplantation, for supporting cell culture of primary tissues, to  
regenerate tissues and in chemotaxis. The polypeptides can also be used  
as a food additive or preservative to increase or decrease storage  
capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
minerals, cofactors and other nutritional components. The present  
invention encodes a novel secreted protein of the invention. Note: The  
sequence data for this patent did not form part of the printed  
Query Match 59.8%; Score 770.4; DB 4; Length 799;  
Best Local Similarity 99.9%; Pred. No. 2.2e-215;  
Matches 771; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
514 GTCTCCAGTGAAGTCTCAACTTTGGCATAAAGAAAAGCGCGGAGTCCAACTGGAATCCT 573  
12 GTCTCCAGTGAAGTCTCAACTTTGGCATAAAGAAAAGCGCGGAGTCCAACTGGAATCCT 71  
574 CAGCATCTTCTGATCTACAGCCCGAGTTTGCCTCGAGCCTCTTCAGACATCAGAGGCG 633  
72 CAGCATCTTCTGATCTTCAAGCCCGAGTTTGCCTCGAGCCTCTTCAGACATCAGAGGCG 131  
634 AATACCTCTGCTGTCTGATCTGGGCCCCACAGAGCGGTATGCAATCTCTAAGTTTACGGGA 693  
132 AATACCTCTGCTGTCTGATCTGGGCCCCACAGAGCGGTATGCAATCTCTAAGTTTACGGGA 191  
694 CTCTCAATGACCACTATCAGTCTACTGGTGGCCCAACGTCAGTACCAGCTTAGGAAAACG 753  
192 CTCTCAATGACCACTATCAGTCTACTGGTGGCCCAACGTCAGTACCAGCTTAGGAAAACG 251  
754 GCGGGGACAAAATTTCTGAAAAAATGAGCAAAAGGCGACCCCATCTTTTATTCAGTGAC 813  
252 GCGGGGACAAAATTTCTGAAAAAATGAGCAAAAGGCGACCCCATCTTTTATTCAGTGAC 911

QY 814 TGTGCTCCAGTTCAGAAACCCCTTTCTACATCATCAGTCACTTAAATCTCACCTGGGT 873  
DB 312 TGTGCTCCAGTTCAGAAACCCCTTTCTACATCATCAGTCACTTAAATCTCACCTGGGT 371  
QY 874 TTCCAAATGAAGGACATGACCCGCTTGTCTAGTGGACCGAAGCAAGGTGAGCAAGAG 933  
DB 372 TTCCAAATGAAGGACATGACCCGCTTGTCTAGTGGACCGAAGCAAGGTGAGCAAGAG 431  
QY 934 ATCTCCCGGGTATCTCGGCTCAGAGGTCTCCAGAAAAAATAAGCTGCCGAGGACACAC 993  
DB 432 ATCTCCCGGGTATCTCGGCTCAGAGGTCTCCAGAAAAAATAAGCTGCCGAGGACACAC 491  
QY 994 GACTCTAAATTCAGAGTGAAGTTGTCTGCGACATTTGTGAGCAAAACATCGGTAAAA 1053  
DB 492 GACTCTAAATTCAGAGTGAAGTTGTCTGCGACATTTGTGAGCAAAACATCGGTAAAA 551  
QY 1054 TTCCACCTTAAGCAAAACGACAGCAAGTCAACCCGAAACCATTCACAGTTTGTAAACAGAC 1113  
DB 552 TTCCACCTTAAGCAAAACGACAGCAAGTCAACCCGAAACCATTCACAGTTTGTAAACAGAC 611  
QY 1114 GTGGATGAAGATAGTCTCTGCGAGGACGAATGCTTAGTTTCCACTTTCAGGCTGGATCC 1173  
DB 612 GTGGATGAAGATAGTCTCTGCGAGGACGAATGCTTAGTTTCCACTTTCAGGCTGGATCC 671  
QY 1174 CTTACACTGAACCCCTTTCTGCTTGACCATCTCTGACATTTGACATTTGAATCACT 1233  
DB 672 CTTACACTGAACCCCTTTCTGCTTGACCATCTCTGCTTGACATTTGAATCACT 731  
QY 1234 CTTCTGACACCTGGCTCTGAGAGACTGCGCAAAAAAAGAAAAA 1285  
DB 732 CTTCTGACACCTGGCTCTGAGAGACTGCGCAAAAAAAGAAAAA 783  
RESULT 8  
ABX73355  
ID ABX73355 standard; DNA; 799 BP.  
AC ABX73355;  
XX  
DT 18-MAR-2003 (first entry)  
XX Human novel polynucleotide #183.  
DE  
XX Human; gene; ds; neural disorder; immune system disorder; renal disorder;  
KW muscular disorder; respiratory disease; reproductive disorder;  
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;  
KW hyperproliferative disorder; inflammatory disease; allergic reaction;  
KW blood related disorder; cancer; immunosuppressive; antinflammatory;  
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;  
KW haemostatic; antiarteriosclerotic.  
XX  
XX Homo sapiens.  
OS  
XX US2002132753-A1.  
PN  
XX  
PD 19-SEP-2002.  
XX  
PF 17-JAN-2001; 2001US-00764864.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 XX (ROSE/) ROSEN C A.  
 PA (RUBE/) RUBEN S M.  
 PA (BARA/) BARASH S C.  
 PI Rosen CA, Ruben SM, Barash SC;  
 XX WPI; 2003-147444/14.  
 DR P-PSDB; ABUS5095.  
 XX  
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,  
 PT inhibiting or preventing e.g. neural, immune system, muscular,  
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or  
 PT renal disorders.  
 XX Claim 1; SEQ ID NO 193; 402pp; English.  
 PS  
 XX The invention relates to human novel polypeptides and their associated  
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene  
 CC therapy for treating, inhibiting or preventing neural disorders, immune  
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis  
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.  
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC (e.g. congenital heart defects, Epstein's anomaly and hypoplastic left  
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage  
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and  
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and  
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood  
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial  
 CC infarction) and cancerous diseases. Sequences ABX71173-ABX74167 represent  
 CC human novel polynucleotides of the invention  
 XX  
 SQ Sequence 799 BP; 247 A; 218 C; 162 G; 172 T; 0 U; 0 Other;  
 Query Match 59.8%; Score 770.4; DB 8; Length 799;  
 Best Local Similarity 99.9%; Pred. NO. 2.2e-215;

	Matches	771;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
QY	514	GTCTCCAGTGAAGTCTCAACTTTTCATATAAAGAAAGCCGGCAGTCCAACTGGGAATCCT	573							
Db	12	GTCTCCAGTGAAGTCTCAACTTTTCATATAAAGAAAGCCGGCAGTCCAACTGGGAATCCT	71							
QY	574	CAGCATCTTCTGATTTTACAAAGCCAGTTTTCGCTCGAGCCTCTTCCAGACATCAGAGGC	633							
Db	72	CAGCATCTTCTGATTTTACAAAGCCAGTTTTCGCTCGAGCCTCTTCCAGACATCAGAGGC	131							
QY	634	AAATACCTGCTGTCTGATCTGGGCCCAACAAGAGCGTATGCAAAATCTCTAAGTTTACGGGA	693							
Db	132	AAATACCTGCTGTCTGATCTGGGCCCAACAAGAGCGTATGCAAAATCTCTAAGTTTACGGGA	191							
QY	694	CTCTCAATGACCACTATCAGTCACTGGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAACG	753							
Db	192	CTCTCAATGACCACTATCAGTCACTGGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAACG	251							
QY	754	GGCGGACAAATTTCTGAAAACATGGACAAAGGCCACCCATCTTTTATTCAGTGCAC	813							
Db	252	GGCGGACAAATTTCTGAAAACATGGACAAAGGCCACCCATCTTTTATTCAGTGCAC	311							
QY	814	TGTGCTCTCCAGTTTTCAGAAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT	873							
Db	312	TGTGCTCTCCAGTTTTCAGAAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGT	371							
QY	874	TTCAAATGAAGACATGACCCCGTGTTCAGTGGACCAAGCAAGAGTGGAGCAAGAG	933							
Db	372	TTCAAATGAAGACATGACCCCGTGTTCAGTGGACCAAGCAAGAGTGGAGCAAGAG	431							
QY	934	ATCTCCCGGTATCGTGGCTCAGAGTCTCAGAAACAATAGTCTGCCGAGAGGACACA	993							
Db	432	ATCTCCCGGTATCGTGGCTCAGAGTCTCAGAAACAATAGTCTGCCGAGAGGACACA	491							
QY	994	GACTCTAAATTTCAAGTGTAAAGTGTCTGCTCGGACATTTGTGAGCAACATCGCGTAAAA	1053							
Db	492	GACTCTAAATTTCAAGTGTAAAGTGTCTGCTCGGACATTTGTGAGCAACATCGCGTAAAA	551							
QY	1054	CTCACCTTAAGCAAAACGACAGCAAGTCAACCGAACAACATTCACAGTTTGTAAACAGAC	1113							
Db	552	CTCACCTTAAGCAAAACGACAGCAAGTCAACCGAACAACATTCACAGTTTGTAAACAGAC	611							
QY	1114	GTGATGAAGATAGCTCTGAGGACGAATGCTTTCAGTCTTCCAGCTTCCAGCTGGATCC	1173							
Db	612	GTGATGAAGATAGCTCTGAGGACGAATGCTTTCAGTCTTCCAGCTTCCAGCTGGATCC	671							
QY	1174	CCTCAGTGAACCTTCTTCTGTTGACCACTCTGCTTCTGACATTTGAACATTCATTGAAC	1233							
Db	672	CCTCAGTGAACCTTCTTCTGTTGACCACTCTGCTTCTGACATTTGAACATTCATTGAAC	731							
QY	1234	CCTCCTGACACCTTGGCTCTGAGAGAGTGCACAAAAAAGAAAAAAGAAAAA 1285								
Db	732	CCTCCTGACACCTTGGCTCTGAGAGAGTGCACAAAAAAGAAAAAAGAAAAA 783								
RESULT 9										
AAI24344										
ID	AAI24344 standard; DNA; 2012 BP.									
XX	AAI24344;									
AC	AAI24344;									
XX	12-OCT-2001 (first entry)									
DT	Probe #14277 for gene expression analysis in human cervical cell sample.									
DB	Probe; human; microarray; gene expression; cervical epithelial cell;									
XX	cervical cancer; 88.									
KW	Homo sapiens.									
OS	WO200157278-A2.									
XX	09-AUG-2001.									



```
Db 1504 TCGGCGCTCGCAACACGCCCCCGGCGCTGCATGCAATCAACCCACTCAGCGCCCTGCAG 1563
Qy 121 TCCGTCCTGAAACAATCAGTGGGCAAGCCACGAGCCCTTGGCGTCACTTCTGCTCC 180
Db 1564 TCCGTCCTGAAACAATCAGTGGGCAAGCCACGAGCCCTTGGCGTCACTTCTGCTCC 1623
Qy 181 AGCCCAAGTTCAGACAAATTCACATGTTCCAGTGTCCACAAGTCCGAATCTCAATGTATGACAAAG 240
Db 1624 AGCCCAAGTTCAGACAAATTCACATGTTCCACAAGTCCGAATCTCAATGTATGACAAAG 1683
Qy 241 CCGGTCCTGAGTCCCTCCACAAAGGTCAGCAGCGGTGTCAGGCGGTACTCTGTTTCAG 300
Db 1684 CCGGTCCTGAGTCCCTCCACAAGTTCAGCAGCGGTGTCAGGCGGTACTCTGTTTCAG 1743
Qy 301 AACACGATCAGCCCAATGACCTGACCAAGTCCAAAAGCAAGACCGAGTCTCTCGCAA 360
Db 1744 AACACGATCAGCCCAATGACCTGACCAAGTCCAAAAGCAAGACCGAGTCTCTCGCAA 1803
Qy 361 GCACAATCTTGATGTCCCACTCAGAACGCTCAGAACGCTCTGTGTGACATCGCGACATGGTC 420
Db 1804 GCACAATCTTGATGTCCCACTCAGAACGCTCAGAACGCTCTGTGTGACATCGCGACATGGTC 1863
Qy 421 AAAGTCTCTCCCAAGACCAACCCCAAGCCAGCTCTCTCCAGGTCCTCCCGCCCATG 480
Db 1864 AAAGTCTCTCCCAAGACCAACCCCAAGCCAGCTCTCTCCAGGTCCTCCCGCCCATG 1923
Qy 481 AAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAG 521
Db 1924 AAGCTGGAATGGATGTGAGGCGCTTTGAGGATGTCTCCAG 1964

RESULT 11
AAI62638
ID AAI62638 standard; DNA; 3604 BP.
XX
AC AAI62638;
XX
DT 19-OCT-2001 (first entry)
DE Human breast or ovarian antigen genomic DNA SEQ ID NO: 288.
KW Human; breast antigen; ovarian antigen; cancer; metastasis; gene therapy;
KW ds.
XX
XX Homo sapiens.
XX WO200155324-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001344.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205513P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
```

08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249279P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-488785/53.  
New isolated nucleic acids and polypeptides, useful for diagnosing,  
treating and/or preventing human diseases and disorders.  
Disclosure; SEQ ID NO 288; 520pp + Sequence Listing; English.  
The present invention provides the protein and coding sequences of a  
number of ovarian and breast antigens. These are shown in AAI62467-  
AAI62572 and AAI62440-AAI62345. The sequences can be used in the  
diagnosis, prevention and treatment of breast and ovarian cancers, and  
their metastases. The present sequence is a genomic sequence of the  
invention. Note: The sequence data for this patent did not form part of  
the printed specification, but was obtained in electronic format directly  
from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
Sequence 3604 BP; 1002 A; 955 C; 853 G; 793 T; 0 U; 1 Other;  
Query Match 30.3%; Score 390.4; DB 4; Length 3604;  
Best Local Similarity 62.4%; Pred. No. 2.1e-103;  
Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;  
52 AGCAATGGGTGGCGCCCTCGCCCAACACGCGCGCCCTGCGCATGCAATCAACCCACTCAGC 111  
1243 AACAACTGGGGATCATCATGGAGCCACTCACCGGAGCCTTCTCTTCATCAACCCGCTGAGC 1302

QY 112 GCGCTGCACTCGTCTCTGAAACAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACCT 171  
DB 1303 GCTTTGCAGTCCATCATGAACACACCCAGCTGGGCAAGGTGTCCAAAGCCCGTG----- 1353  
QY 172 TCCTGCTCAGCCCAAGTTCAAGCAATTTCCATGTTCCACAAGTCGATCTCAATGTC 231  
DB 1354 -----AGTCCCTCGCTGGACCCGCTGGCGATGCTGTACAAAGATCAGCAACAGCATG 1404  
QY 232 ATGGACAAGCCGGTCTTGAGTCTGCTCCCAACAAGGTGAGCCAGCGGTGTCCAGGCGCTAC 291  
DB 1405 CTGGACAAGCCGGTGTATCCCGCCACCCCTGTGAAGCAGGCGCGATGCCATCGACCGCTAC 1464  
QY 292 CTGTTTGAAGACAGCGATCAGCCCATTCAGCTGACCAAGTCCAAAGCAAGAAAGCCGAG 351  
DB 1465 TATTATGAAAACAGCGACCGCCCATTCGATTAACTCAAGTCCAAAGCAAGCCGCTGGTG 1524  
QY 352 TCCTGCAAGCAATCTTTGATGTCCCACTCAGAGCAGCGCTCTGCTGACATCGCC 411  
DB 1525 TCCAGCGTGGTGTATTCGGTGGCATCAGCTCTGCGGAGAGCGCACTCATGACATCTCC 1584  
QY 412 GACATGGTCAAGTCTCTCCCAAGCCACCCCAAGCCAGCGCTCTCTCCAGGGTCT 471  
DB 1585 GACATGGTGAAGAACTCAGAGCGCGCTGAGCGCCCAAGTCTCCAGCGCTCCACAGTT 1644  
QY 472 CCCCCCATGAGCTGGAATGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531  
DB 1645 TCAGA--GAAGTCCGATGCTGATGCGAGCAGCTTTGAGGAGCGTGGAGAGTGTCA 1701  
QY 532 ACTTTGCAATAAAGAAAGCCGCGAGTCCAACTCGAATCTCTCAGCATCTTCTGATTCTA 591  
DB 1702 CCGGTCCACAAGAGGAAGGCGCGCAGTCCAACTGGAACCCCGCAGCACCTTCTCATCTCG 1761  
QY 592 CAAAGCCAGTTTGGCTCGAGCGCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGTGAT 651  
DB 1762 CAGGCCAGTTCGCTCGAGCTTGGGGAGACCAAGAGGGCAAGTACATCATGTGCGAC 1821  
QY 652 CTGGGCCCCACAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCACTATC 711  
DB 1822 TTGGGCCCCAGAGAGGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACTC 1881  
QY 712 AGTCATGCTGGCCAAAGTCAAGTACAGCTTAGAANAACGGGCGGGAACAAATTTCTG 771  
DB 1882 AGCCACTGGCTGGCCAAATGTGAAGTACCAGTTGAGGAGACAGGGGGAAACGAAATTCCTA 1941  
QY 772 AAAAAATGGAACAAAGGCCACCCCATCTTTTATTCAGTGAAGTGTGCTCCCTCCAGTTCAGA 831  
DB 1942 AAGAACCTGGACACAGGGCATCTGTTTCTTTTGAACAGATGTGCTCTCAGTTCAGA 2001  
QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATG 891  
DB 2002 ACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTGGGCTTCAAGCTGAAGGATCTC 2061  
QY 892 ACCCGTTGCTAGTGGACCAAGCAAGGTGGAGCAAGATCTCCCGGGTATCGTCG 951  
DB 2062 TCCAAAGTCTCCACT---CAATCAGATTCAAGAACACAGCAAAATGTTTCGAAAGCTCCACC 2118  
QY 952 GCTCAGAGTCTCCAGAAACAATAGTGTGCGGAGGAGACACAGACTCTAAATTCAGATGT 1011  
DB 2119 AACAAACTCTCTGGGCCCATCTGGGGGCCACCGAGGAAGATTTGGGCTCCCATTCATGT 2178  
QY 1012 AAGTTGTGCTGTGGGACATTTGTGAGCAAAACATGCGGTAAAACTCCACCTTAAGCAAAACG 1071  
DB 2179 AAGCTCTCAACCGGACTTTTGGGAGCAAGCAGCAGTCAAACTGACCTTAGTAAGACC 2238  
QY 1072 CACAGCAAGTCCCGGAACACCAATTCAGATTTGTAACAGAGCTGATGAAGATAGC 1129  
DB 2239 CACGGCAAGTCTCCCGAGGACCACCTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 2296  
RESULT 12  
AAL06732  
ID AAL06732 standard; DNA; 3604 BP.

XX AAL06732;  
AC 21-NOV-2001 (first entry)  
DT Human reproductive system related antigen DNA SEQ ID NO: 9420.  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW cancer; gene therapy; ds.  
KW Homo sapiens.  
OS WO200153320-A2.  
PW 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001339.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227003P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229503P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 13-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.

PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
XX used in preventing, treating or ameliorating a medical condition.  
XX  
XX Disclosure; SEQ ID NO 9420; 1297pp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
XX number of human reproductive system related antigens. These can be used  
XX in the prevention and treatment of reproductive system disorders,  
XX including cancer. The present sequence is a genomic sequence encoding a  
XX protein of the invention  
XX  
XX Sequence 3604 BP; 1002 A; 955 C; 853 G; 793 T; 0 U; 1 Other;  
XX  
XX  
XX Query Match 30.3%; Score 390.4; DB 4; Length 3604;  
XX Best Local Similarity 62.4%; Pred. No. 2.1e-103;  
XX Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;  
XX  
QY 52 AGCAATGGTGGCGCTCCGCCAACACGCGCCGCTGCCATGATCAACCCACTCAGC 111  
DB 1243 AACAACCTGGGATCATATGACCACTCACCAGGACCTTCTTCAACCCGCTGAGC 1302  
QY 112 GGCCTGCACTCGTCTGAAACATCACTTGGGCAAGCCACGAGCCCTTGGGCTCACCT 171  
DB 1303 GCTTTGCACTCCATCAATGAACCCACCTGGGCAAGGTGTCCAAGCCCGTG----- 1353  
QY 172 TCGTGTCCAGCCCAAGTTCAAGCACAATTCATGTTCCATGATCAAGTCGAATCTCAATGTC 231  
DB 1354 -----AGTCCCTCGCTGGACCCGCTGGCGATGCTGTACAAGATCAGCAACAGCATG 1404  
QY 232 ATGGACAAGCCGCTCTTGAGTCTGCTCCACAAGTTCAGCCAGCGTGTCCAGGCGCTAC 291  
DB 1405 CTGGACAGCCGCTGTACCCGCCACCTCTGTGAACAGCGCGATGCCATGACCGCTAC 1464  
QY 292 CTGTTTGAACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAAAGCCGAG 351  
DB 1465 TATTATGAACACAGCGACCGACCCATTGACTTAACCAAGTCCAAAGAACAGCGCTGGTG 1524  
QY 352 TCGTCCGAAGCAAACTTTGTATGTCTCCCACTCAGAAAGCAGCTCTGTGATCAGCTCGCC 411  
DB 1525 TCCAGCGTGGCTGATTCGGTGGCATCACCTCTCGGGGAGAGCGCACTCATGACATCTCC 1584  
QY 412 GACATGTCAAAGTCTCCCAAGACCAACCCCAAGCCAGCCCTCTCTCCAGGTC 471  
DB 1585 GACATGTTGAAACCTCTCAGGCCGCTGACGCCCAAGTCTCCAGCCCTCCACAGTT 1644  
QY 472 CCCCCATGAAGCTGGAATGGAATGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531  
DB 1645 TCAGA---GAAGTCCGATGCTGATGCGCAGCAGCTTTGAGGAGCGTTGGACGAGTGCA 1701  
QY 532 ACTTTCATAAAGAAAGCCCGGAGTCCCACTGGAATCTCAGCATCTTCTGATCTTA 591  
DB 1702 CCGGTCCACAAGAGGAGGCGCGCAGTCCCACTGGAACCCCGCAGCACCTTCTCATCTG 1761  
QY 592 CAAAGCCAGTTTCCCTCGAGCTCTTCCAGACATCAGAGGCAAAATACCTGCTGTGAT 651  
DB 1762 CAGGCCAGTTTCGCTTCGAGCTTGGGGAGACCAAGGAGGCAAGTACATCATGTCGAGC 1821  
QY 652 CTGGGCCCAAGAGCGTATGCAAAATCTCTAAGTTTACGGGACTCTCAATGACCACATC 711

DB 1822 TTGGCCCGCAGGAGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACCATC 1881  
QY 712 AGTCACTGCTGGCCAAAGTCAAGTACCAGCTTAGGAAAAACGGGGGACAAAATTTCTG 771  
DB 1882 AGCCACTGGCTGGCCAAATGTGAAGTACCAGTTGAGAGGACAGGGGGAACGAATTCCTA 1941  
QY 772 AAAAATCATGGACAAAGGCCACCCCATCTTTTATTGCACTGAGTGTGCTCTCCAGTTCA 831  
DB 1942 AAGAACCTGGACACAGGGCATCTCTGTTTCTTTTCAACGATTGTGCTCTCAGTTCA 2001  
QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGACATG 891  
DB 2002 ACTGCTTCTACATACATAAGTCAATTTGGAGACACACTTGGGCTTCAAGGATCTC 2061  
QY 892 ACCCGCTTGTGAGTGGACCAAGCAAGGTGGAGGACAGAGATCTCCCGGTATCGTCG 951  
DB 2062 TCCAAAGTGCCTACT---CAATCAGATTCAAGAACAGCAGATGTTTCGAAAGTCTCACC 2118  
QY 952 GCTCAGAGTCTCCAGAAACATAGTCTGCCGAAGAGGACACAGACTCTAAATTCAGTGT 1011  
DB 2119 AACAAATCTCTGGGCCACTGGGGCCACCGAGGAAGACTTGGGCTCCCAATTCATGT 2178  
QY 1012 AAGTTGTGCTGCGACATTTGTGAGCAAAACATGCGGTAAAACTCCACCTAAGCAAAAG 1071  
DB 2179 AAGTCTGCAACCGGACTTTTTCGAGCAAGCAGCAGTCAAACTGCACCTTAGTAAGAC 2238  
QY 1072 CACAGCAAGTCAACCCGACCACTTTCACAGTTTGTAAACAGACGTGATCAAGATAGC 1129  
DB 2239 CACGCAAGTCTCCCGAGGACCACCTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 2296  
XX  
XX RESULT 13  
XX ABK34640  
XX ID ABK34640 standard; cDNA; 3907 BP.  
XX AC ABK34640;  
XX DT 08-MAY-2002 (first entry)  
XX DE Human cDNA for novel secreted protein, SEQ ID 409.  
XX  
XX Human; ss; gene; secreted protein; immune deficiency; viral infection;  
XX bacterial infection; fungal infection; autoimmune disorder; burn;  
XX rheumatoid arthritis; multiple sclerosis; autoimmune thyroiditis;  
XX diabetes; allergy; asthma; neurodegenerative disease; Crohn's disease;  
XX Alzheimer's disease; Parkinson's disease; liver fibrosis; tumour;  
XX coagulation disorder; haemophilia; inflammatory disorder; ulcer;  
XX tissue regeneration; wound healing; haematopoiesis; myeloid deficiency;  
XX lymphoid cell deficiency.  
XX  
XX Homo sapiens.  
XX  
XX WO200177290-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 29-MAR-2001; 2001WO-US010295.  
XX  
XX 06-APR-2000; 2000US-0194941P.  
XX  
XX (GEMY ) GENETICS INST INC.  
XX  
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;  
XX Gulukota K, Graham JR;  
XX  
XX WPI; 2002-179323/23.  
XX  
XX Six hundred and twenty five polynucleotides derived from a variety of  
XX human tissue sources which encode secreted proteins, useful for treating  
XX immune deficiencies and disorders such as autoimmune disorders.  
XX  
XX Claim 1; Page 210-211; 339pp; English.





Query Match		30.3%;	Score 390.4;	DB 10;	Length 7081;
Best Local Similarity		62.4%;	Pred. No. 3e-103;		
Matches 673;		Conservative 0;	Mismatches 381;	Indels 24;	Gaps 3;
QY	52	AGCAATGGTGGCCCTCGCCACACACGCGCCGCTTCCATGTCATCAACCCACTCAGC	111		
DB	4726	AACAACCTGGGATCATCTGGACCACTCACCGAGCCTTCTTCAACCGCTGAGC	4785		
QY	112	GCCTGCACTGCTCTGAAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACT	171		
DB	4786	GCTTTGCACTGCTCAATGAACACCACTGGCAAGGTGTCCAAGCCCGTG-----	4836		
QY	172	TCCTGCTCAGCCCAAGTTCAAGCAAAATTTCCATGTTCCACAAGTCGAATCTCAATGTC	231		
DB	4837	-----AGTCCCTCGCTGGACCCGCTGGCGATGCTGTACAGATCAGCAACAGCATG	4887		
QY	232	ATGGACAAGCCGCTTGTAGTCTGCTCCACAAAGTCAGCCAGCGTGTCCAGGCGCTAC	291		
DB	4888	CTGGACAAGCCGCTGTACCCGCGCACCCCTGTGTAAGCAGGCGCATGCCATCGACTAC	4947		
QY	292	CTGTTTGAGAACAGCATCAGCCATGACCTGACCAAGTCACCAAGTCCAAAGCAAGAAAGCCGAG	351		
DB	4948	TATTATGAAGAACAGCAGCAGCCCATTTGACTTAACCAAGTTCGAAGCAAGCGCGTGTG	5007		
QY	352	TCCTCGCAAGCACAATCTTTGTATGTCCTCCACCTCAGAAGCAGCGCTCTGTCTGACATCGCC	411		
DB	5008	TCAGCGCTGCTGATTTGGTGGCATCACCTCTCGGGAGAGCGCATCTATGACATCTCC	5067		
QY	412	GACATGGTCAAAAGTCTCTCCCAAGACCAACCCCAAGCCAGCCTCTCTCCAGGGTC	471		
DB	5068	GACATGGTGAAGAACCTCAGCGCCGCTGAGCGCCCAAGTCTCTCCAGCCCTCCACAGTT	5127		
QY	472	CCCCCATGAAGCTGAATGATGTGAGCGCTTTGAGATGTCTCCAGTGAAGTCTCA	531		
DB	5128	TCAGA---GAAAGTCCGATGCTGATGAGCAGCAGCTTTGAGGAGGCGTTGAGCAGGCTGCA	5184		
QY	532	ACTTTGCAATAAAGAAAGCGCGGAGTCCAATCGAATCTCTCAGCATCTTCTGATTCTA	591		
DB	5185	CGGTTCCACAGAGAGAGAGCGCGGAGTCCAACTGGAAACCGGAGCAGCTTCTCATCTG	5244		
QY	592	CAAGCCAGTTTGCCTCGAGCCTCTTTCCAGACATCAGAGGGCAATACCTGTCTGTAT	651		
DB	5245	CAGGCCAGTTTGCCTCGAGCTTGGCGGAGACACAGAGGGCAAGTACATCATGTGGAC	5304		
QY	652	CTGGGCCCCAAGAGCGGTATGAAATCTTAAGTTTACGGGACTCTCAATGACCATATC	711		
DB	5305	TTGGGCCCCGAGGAGGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCCACATC	5364		
QY	712	AGTCACTGGCTGGCCCAAGTCAAGTACCAGCTTAGGAAACGGGCGGACAAATTTCTG	771		
DB	5365	AGCCACTGGCTGGCCAAATGTGAATACCAAGTTGAGAGGACAGGGGGAAGAAATTCCTA	5424		
QY	772	AAAAATGGAACAAAGCCACCCCATCTTTTATTCAGTGACTGTGCTCCCAAGTTTCA	831		
DB	5425	AAGAACTGGACACAGGCGATCTGTTTCTTTTGAACAGATTGTGCTCTCAAGTTTCA	5484		
QY	832	ACCCCTTCTACCTACATCAGTCACTTAGAATCTCACCCTGGGTTTCCAAATGAAGCATG	891		
DB	5485	ACTGCTTCTACATACATAAGTCAATTTGGAGACACCTTGGGCTTACGCTGGAAGGATCTC	5544		
QY	892	ACCCGCTTGTGCTGGACCAAGCAAGGTGGACAGAGATCTCCCGGGTATCGTGC	951		
DB	5545	TCCAAGTGCACCT---CAATGATTCAGAACAGCAGAGATGTTTGGAAAGTCTCCACC	5601		
QY	952	GCTCAGAGGTCTCCAGAAACAATAGTGGCGAAGAGACACAGACTCTAAATTCAGTGT	1011		
DB	5602	AACAAAACCTCTGGGCCCACTGGGGGCCACCGAGGAAGACTTGGGCTCCCAATTCCAATGT	5661		
QY	1012	AAGTTGTGCTCGGACATTTGTAGCAACATGCGGTAAACTCCACCTTAACAAACG	1071		
DB	5662	AAGCTCTGCAACCGGACTTTTGGAGCAAGCAGCAGCTCAACTGACCTTAGTAAGACC	5721		

QY	1072	CACAGCAAGTCACCGAACACCACTTACAGTTTGTAAACAGACGTGATGAAGAATAGC	1129
DB	5722	CACGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGTGAGAAACAGTAGC	5779
RESULT 15			
ID	ADC31952		
XX	ADC31952 standard; cDNA; 1079 BP.		
AC	ADC31952;		
DT	18-DEC-2003 (first entry)		
XX	Human novel cDNA contig sequence, SEQ ID NO:2034.		
DE	Human; diagnostic; drug screening; forensics; gene mapping;		
XX	biodiversity assessment; Parkinson's disease; Alzheimer's disease;		
KW	neurodegenerative diseases; anaemia; platelet disorder; wound; burns;		
KW	ulcers; osteoporosis; autoimmune disease; cancer;		
KW	molecular weight marker; food supplement; antiparkinsonian; nootropic;		
KW	neuroprotective; anti-anaemic; anticoagulant; thrombolytic; vulnery;		
KW	antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;		
KW	gene therapy; chromosome 20; ss.		
XX	Homo sapiens.		
OS	WO2003029271-A2.		
XX	10-APR-2003.		
PD	24-SEP-2002; 2002WO-US030474.		
PF	24-SEP-2001; 2001US-0324631P.		
XX	(HYSE-) HYSEQ INC.		
PA	Tang TY, Zhang J, Ren P, Xue AJ, Zhao QA, Wang J, Wehrman T;		
PI	Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;		
PI	Haley-Vicente D, Drmanac RT;		
XX	WPI; 2003-371981/35.		
DR	P-PSDB; ADC32719.		
XX	New polynucleotide and polypeptide useful for diagnosing, preventing or		
PT	treating conditions such as neurodegenerative diseases, anemias, platelet		
PT	disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or		
PT	cancer.		
XX	Example 2; SEQ ID NO 2034; 1185pp; English.		
PS	The invention relates to 971 novel human cDNA sequences (ADC29919-		
XX	ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The		
CC	invention also relates to nucleic acid sequences over 99% identical with		
CC	the novel human cDNAs. The invention additionally encompasses expression		
CC	vectors and host cells comprising a nucleic acid of the invention; the		
CC	recombinant production of a polypeptide of the invention; an antibody		
CC	against a polypeptide of the invention; a method of detecting		
CC	polynucleotides or polypeptides of the invention; and methods of		
CC	identifying a compound which binds to a polypeptide of the invention. The		
CC	invention further discloses methods of patenting, treating or		
CC	ameliorating a medical condition; kits comprising polynucleotide probes		
CC	and/or monoclonal antibodies for carrying out the methods of the		
CC	invention; methods for the identification of compounds that modulate the		
CC	expression or activity of the polynucleotide and/or polypeptide; and 767		
CC	contig sequences corresponding to the cDNA sequences of the invention		
CC	(ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628		
CC	-ADC33394). The nucleic acids and polypeptides of the invention are		
CC	useful in diagnostics, drug screening, forensics, gene mapping, in the		
CC	identification of mutations responsible for genetic disorders or other		
CC	traits, for assessing biodiversity, and in producing many other types of		
CC	data and products dependent on DNA and amino acid sequences. They are		
CC	also used for treating diseases such as Parkinson's disease, Alzheimer's		
CC	disease and other neurodegenerative diseases, anaemia, platelet		

CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or  
CC cancer. The nucleic acids may also be used as hybridisation probes or  
CC primers, and in the recombinant production of a protein. The polypeptides  
CC are also useful in generating antibodies, as molecular weight markers,  
CC and as food supplements. The present sequence represents a human contig  
CC sequence used in an example of the invention. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1079 BP; 322 A; 317 C; 257 G; 183 T; 0 U; 0 Other;

Query Match 30.3%; Score 390; DB 10; Length 1079;  
Best Local Similarity 100.0%; Pred. No. 1.5e-103;  
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GAGGCGCAGCGAAGAGAGAAACCCAGCCGCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG	60
Db	617	GAGGCGCAGCGAAGAGAGAAACCCAGCCGCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG	676
QY	61	TGCGCCCTTGCCCAACACGCGCCCGCCCTGGCCATCATCAACCCACTCAGCGCCCTGCAG	120
Db	677	TGCGCCCTTGCCCAACACGCGCCCGCCCTGGCCATCATCAACCCACTCAGCGCCCTGCAG	736
QY	121	TCCGTCTTGAAACAATCACTTGGGGCAAGCCACGGAGCCCTTGGCTCACCCTTCCTGCTCC	180
Db	737	TCCGTCTTGAAACAATCACTTGGGGCAAGCCACGGAGCCCTTGGCTCACCCTTCCTGCTCC	796
QY	181	AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCAATGACAAG	240
Db	797	AGCCCAAGTTCAAGCACAAATTTCCATGTTCCACAAGTCGAATCTCAATGTCAATGACAAG	856
QY	241	CCGGTCTTGAGTCTCTGCCCTCCACAAGGTCAGCCAGCGGTGCCAGGCGCTACCTGTTTGAG	300
Db	857	CCGGTCTTGAGTCTCTGCCCTCCACAAGGTCAGCCAGCGGTGCCAGGCGCTACCTGTTTGAG	916
QY	301	AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGACCGAGTCCTCGCAA	360
Db	917	AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGACCGAGTCCTCGCAA	976
QY	361	GCACAATCTTGATGTCCCCACCTCAGAAG	390
Db	977	GCACAATCTTGATGTCCCCACCTCAGAAG	1006

Search completed: May 5, 2005, 00:37:17  
Job time : 901.572 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:46:20 ; Search time 289.356 Seconds  
(without alignments)  
7283.499 Million cell updates/sec

Title: US-08-731-499-5  
Perfect score: 1288  
Sequence: 1 GAGGCACCGAGAGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY\_NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1288	100.0	1288	2	US-08-680-395-5
2	1288	100.0	1288	4	US-08-892-695-5
3	377.8	29.3	2918	4	US-09-510-543-21
4	256.4	19.9	1144	2	US-08-828-008-1
5	113	8.8	282	2	US-08-828-008-3
6	41.6	3.2	258	4	US-09-023-655-773
7	40	3.1	31623	4	US-09-949-016-15945
8	37.6	2.9	601	4	US-09-949-016-115533
9	37.6	2.9	601	4	US-09-949-016-115626
10	37.6	2.9	601	4	US-09-949-016-115719
11	37.6	2.9	601	4	US-09-949-016-115812
12	37.6	2.9	601	4	US-09-949-016-115905
13	37.6	2.9	601	4	US-09-949-016-115998
14	37.6	2.9	601	4	US-09-949-016-116091
15	37.6	2.9	601	4	US-09-949-016-116184
16	37.6	2.9	601	4	US-09-949-016-116277
17	37.6	2.9	601	4	US-09-949-016-116370
18	37.6	2.9	110266	4	US-09-949-016-14913
19	37.6	2.9	110266	4	US-09-949-016-14914
20	37.6	2.9	110266	4	US-09-949-016-14915
21	37.6	2.9	110266	4	US-09-949-016-14916
22	37.6	2.9	110266	4	US-09-949-016-14917
23	37.6	2.9	110266	4	US-09-949-016-14918
24	37.6	2.9	110266	4	US-09-949-016-14919
25	37.6	2.9	110266	4	US-09-949-016-14920
26	37.6	2.9	110266	4	US-09-949-016-14921
27	37.6	2.9	110266	4	US-09-949-016-14922

C 28	37	2.9	505	4	US-09-621-976-15639	Sequence 15639, A
C 29	37	2.9	1389	4	US-09-902-540-3663	Sequence 3663, Ap
C 30	37	2.9	23847	4	US-09-902-540-1177	Sequence 1177, Ap
C 31	36.8	2.9	118136	4	US-09-949-016-12439	Sequence 12439, A
C 32	36.2	2.8	399	4	US-09-621-976-8976	Sequence 8976, Ap
C 33	36.2	2.8	1497	4	US-09-220-132-94	Sequence 94, Appl
C 34	35.8	2.8	1905	4	US-09-902-540-2711	Sequence 2711, Ap
C 35	35.8	2.8	14861	4	US-09-902-540-1127	Sequence 1127, Ap
C 36	35.4	2.7	1141	4	US-09-889-463A-37	Sequence 37, Appl
C 37	35.4	2.7	7007	4	US-09-949-016-12430	Sequence 12430, A
C 38	35.4	2.7	7007	4	US-09-949-016-15989	Sequence 15989, A
C 39	35.4	2.7	50836	4	US-09-949-016-16722	Sequence 16722, A
C 40	35.2	2.7	2406	4	US-09-902-540-5545	Sequence 5545, Ap
C 41	35.2	2.7	11812	4	US-09-902-540-1041	Sequence 1041, Ap
C 42	35.2	2.7	26452	4	US-09-949-016-15822	Sequence 15822, A
C 43	35.2	2.7	64190	4	US-09-949-016-14712	Sequence 14712, A
C 44	35.2	2.7	64190	4	US-09-949-016-14713	Sequence 14713, A
C 45	35.2	2.7	133613	4	US-09-949-016-15824	Sequence 15824, A

ALIGNMENTS

RESULT 1  
US-08-680-395-5  
; Sequence 5, Application US/08680395  
; Patent No. 5892010  
; GENERAL INFORMATION:  
; APPLICANT: Gray, Joe W.  
; APPLICANT: Collins, Colin  
; APPLICANT: Hwang, Soo-in  
; APPLICANT: Godfrey, Tony  
; APPLICANT: Kowbel, David  
; APPLICANT: Rommens, Johanna  
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their  
; USES  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680.395  
; FILING DATE: 15-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-0689000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1288 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; FEATURE:  
; NAME/KEY: -  
; LOCATION: 1..1288 /note= "cdna clone 41.1 with homology  
; OTHER INFORMATION: /note= "cdna clone 41.1 with homology  
; OTHER INFORMATION: to homeobox T shirt gene from  
; OTHER INFORMATION: Drosophila"



241 CCGGTCCTTGAAGTCCTCCACAGGTCAGCCAGGTCCTCAGGCGCTACCTGTTGAG 300  
Db |||  
241 CCGGTCCTTGAAGTCCTCCACAGGTCAGCCAGGTCCTCAGGCGCTACCTGTTGAG 300  
Qy |||  
301 AACACGGATCAGCCATTCAGCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCCTCGAA 360  
Db |||  
301 AACACGGATCAGCCATTCAGCTGACCAAGTCCAAAAGCAAGAAAGCCGAGTCCTCGAA 360  
Qy |||  
361 GCACAACTTGTATGTCTCCACCTCAGAGCAAGCTCTGTCTGATCGACATCGCGACATGTC 420  
Db |||  
361 GCACAACTTGTATGTCTCCACCTCAGAGCAAGCTCTGTCTGATCGACATCGCGACATGTC 420  
Qy |||  
421 AAAGTCCTCCCAAGACCAACCCCAAGCCAGCTCTCTCTCCAGGTCCTCCCAAGTC 480  
Db |||  
421 AAAGTCCTCCCAAGACCAACCCCAAGCCAGCTCTCTCTCCAGGTCCTCCCAAGTC 480  
Qy |||  
481 AAGCTGGAAATGGATGTGAGGCGCTTTCAGGATGTCTCAGTGAAGTCTCAACTTTGAT 540  
Db |||  
481 AAGCTGGAAATGGATGTGAGGCGCTTTCAGGATGTCTCAGTGAAGTCTCAACTTTGAT 540  
Qy |||  
541 AAAGAAAGGCGGAGTCCTCAACTGGAATCTCAGCATCTTCTGATCTCAAGCCCGAG 600  
Db |||  
541 AAAGAAAGGCGGAGTCCTCAACTGGAATCTCAGCATCTTCTGATCTCAAGCCCGAG 600  
Qy |||  
601 TTTGGCTCGAGCTCTCCAGACATCAGAGGCAATACCTCTCTGATCTGGGCCA 660  
Db |||  
601 TTTGGCTCGAGCTCTCCAGACATCAGAGGCAATACCTCTCTGATCTGGGCCA 660  
Qy |||  
661 CAAGAGCGTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720  
Db |||  
661 CAAGAGCGTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720  
Qy |||  
721 CTGGCCAAACGTCAGTCAAGTTCAGGAAACCGGCGGACAAAATTTCTGAAAACATG 780  
Db |||  
721 CTGGCCAAACGTCAGTCAAGTTCAGGAAACCGGCGGACAAAATTTCTGAAAACATG 780  
Qy |||  
781 GACAAAGGCCACCCATCTTTTATGTCAGTACTGTGCTCCAGTTCAGAACCCCTTCT 840  
Db |||  
781 GACAAAGGCCACCCATCTTTTATGTCAGTACTGTGCTCCAGTTCAGAACCCCTTCT 840  
Qy |||  
841 ACCTACATCAGTCACCTAGAAATCTCAGCTGGGTTTCCAAATGAAGGACATGACCGCTTG 900  
Db |||  
841 ACCTACATCAGTCACCTAGAAATCTCAGCTGGGTTTCCAAATGAAGGACATGACCGCTTG 900  
Qy |||  
901 TCAGTGGACCAAGCAAGGTCGAGCAAGAGATCTCCCGGTTATCGTGGCTCAGAGG 960  
Db |||  
901 TCAGTGGACCAAGCAAGGTCGAGCAAGAGATCTCCCGGTTATCGTGGCTCAGAGG 960  
Qy |||  
961 TCTCAGAAACATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGTGAAGTTGTGC 1020  
Db |||  
961 TCTCAGAAACATAGCTGCCGAAGAGGACACAGACTCTAAATTCAGTGAAGTTGTGC 1020  
Qy |||  
1021 TGTCCGACATTTGTGAGCAACATCGGTAAACCTCCACCTAAGCAAAACGACAGCAAG 1080  
Db |||  
1021 TGTCCGACATTTGTGAGCAACATCGGTAAACCTCCACCTAAGCAAAACGACAGCAAG 1080  
Qy |||  
1081 TCACCCGAACACCATTCACAGTTTGAACAGAGTGGATGAAGATAGCTCTGAGGAGC 1140  
Db |||  
1081 TCACCCGAACACCATTCACAGTTTGAACAGAGTGGATGAAGATAGCTCTGAGGAGC 1140  
Qy |||  
1141 AATGCTTAGTTTCCATTTTCAGGCTGGATCCCTCAGACTGAACCCCTTCTGCTTGA 1200  
Db |||  
1141 AATGCTTAGTTTCCATTTTCAGGCTGGATCCCTCAGACTGAACCCCTTCTGCTTGA 1200  
Qy |||  
1201 CCATCTGCTCTGACATTTGAATGAACTCTCTCTGACACCCCTGCTCTGAGAGA 1260  
Db |||  
1201 CCATCTGCTCTGACATTTGAATGAACTCTCTCTGACACCCCTGCTCTGAGAGA 1260  
Qy |||  
1261 CTGCCAAAAAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1288  
Db |||  
1261 CTGCCAAAAAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTTCAATTTT 1288

## RESULT 3

US-09-510-543-21  
; Sequence 21, Application US/09510543  
; Patent No. 6517837

## GENERAL INFORMATION:

; APPLICANT: Scanlan, Matthew J.  
; APPLICANT: Chen, Yao-Tsang  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Old, Lloyd J.

; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES ASSOCIATED WITH COLON CANCER AND  
; FILE REFERENCE: LUD-5506.1-JEL/NDH

; CURRENT APPLICATION NUMBER: US/09/510,543

; CURRENT FILING DATE: 2000-02-22

; PRIOR APPLICATION NUMBER: US 09/102,322

; NUMBER OF SEQ ID NOS: 22

; SEQ ID NO 21

; LENGTH: 2918

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-510-543-21

Query Match 29.3%; Score 377.8; DB 4; Length 2918;

Best Local Similarity 62.3%; Pred. No. 1.8e-109;

Matches 672; Conservative 0; Mismatches 382; Indels 25; Gaps 4;

Qy 52 AGCAATGGGTGGCGCTCTGCCAACACCGCCCGGCTTGCATGTCATCAACCCACTCAGC 111

Db 1429 AACAACTCTGGGATCATCATGGACCACTCACGGAGCTTCTTCTCATCAACCGCTGAGC 1488

Qy 112 GCCCTGAGTCGCTCTGAAACATCATTTGGGCAAGCCACGGAGCCCTTGGCGTCACT 171

Db 1489 GCTTTCAGTTCATCATGAACACCACTGGGCAAGGTGTCCAAGCCCGTG----- 1539

Qy 172 TCCTGCTCCAGCCCAAGTTCAAGCACAAATTTCCATGTTTCCCAAGTCGAATCTCAATGTC 231

Db 1540 -----AGTCCCTCGCTGGACCCGCTGGGATGCTGTACAGATCAGCAACAGCATG 1590

Qy 232 ATGGAACAGCGGCTTTCAGTCTTCCCTCCACAGGTCAGCCAGCGTGTCTCAGGCGCTAC 291

Db 1591 CTGGACAGCGGCTGTACCCCGCACCTCTGTGAAGAGCGGCGATGCCATCGCCGTAC 1650

Qy 292 CTGTTTGAACAGCGATCAGCCATTCGACCTGACCAAGTTCAAAAGCAAGAAAGCCGAG 351

Db 1651 TATTATGAAAACAGCGACAGCCCATTTGACTTAACCAAGTCCAAAGCAAGAGCCGCTGTG 1710

Qy 352 TCCTCGCAAGCACAAATCTTGTATGTCCTCCACCTCAGAGCAGCTCTGTCTGACATCGCC 411

Db 1711 TCCAGCGTGGCTGATTCGGTGGCATCACCTCTGCGGGAGAGCGCACTCATGACATCTCC 1770

Qy 412 GACATGGTCAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGCTC 471

Db 1771 GACATGGTGAAGAACTTCAGAGCGCTGAGCGCCCAAGTCTCTCCAGCCCTCCAGT 1830

Qy 472 CCCCCATGAAGCTGGAATGATGTGAGGCGCTTTTGGAGTGTCTCCAGTGAAGTCTCA 531

Db 1831 TCAGA---GAAGTCGATGCTGATGGCAGAGCTTTGAGGAGCGCTTGGACAGCTGTCA 1887

Qy 532 ACTTGTCAATAAAGAAAGGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCA 591

Db 1888 CCGGTCCCAAGAGAAAGGCGGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCA 1947

Qy 592 CAAGCCCATGTTGCTCGAGCTCTTCCAGCATCAGAGGGCAAAATACCTGCTGTCTGAT 651

Db 1948 CAGGCCAGTTCGCTTCAGCTTGGGGAGACCAAGAGGCAAGTACATCATGTCGAG 2007

Qy 652 CTGGGCCCAAGAGCGTATGCAAACTCTTAAGTTTACGGAGTCTCAATGACCACTATC 711

Db 2008 TTGGGCCCGAGAGAGGCTGACATCTCGAAGTTTACTGGGCTCTCCATGACCACTATC 2067

Qy 712 AGTCACTGGTGGCCAAAGTCAAGTACAGCTTACGAAAACGGGCGGACAAATTTCTG 771

Db 2068 AGCCACTGGCTGCCCAATGTGAAGTACCAGTTGAGGAGGACAGGGGAAACGAAATTCCTA 2127  
Qy 772 AAAACATGACAAAGCCACCCCATCTTTTATTGCAGTGACTGTGCTCCCGATTTCAGA 831  
Db 2128 AGAACTGGACACAGGGCATCTGTTTCTTTTGAAGATGTGCGCTCTCAGTTTCCA 2187  
Qy 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGACATG 891  
Db 2188 ACTGCTTCTACATACATAGTCATTTGGAGACACACTTGGCTTCAGCTGAAGATCTC 2247  
Qy 892 ACCGCTTCTGAGTGACACAGAAAGCAAGTGGAGCAAGATCTCCCGGGTATCGTCG 951  
Db 2248 TCAAAGTCCCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCGAAAGTCTCTCACC 2304  
Qy 952 GCTCAGAGTCTCCAGAAACATAGCTGCCGAGAGGACACAGACTCTAAATTCAGTGT 1011  
Db 2305 ACAAAACTCTGGGCCCATCTGGGGCCACCGAGGAAGACTTGGCTTCCACATTCCAATGT 2364  
Qy 1012 AAGTTGTGCTGGGACATTTGTG-AGCAAAACATGCGGTAAAACTCCACTAAGCAAAAC 1070  
Db 2365 AAGCTCTGCAACGGGACTTTTGGAGAGCAAGCAGCAGTCAAACTGCACCTTAGTAAGAC 2424  
Qy 1071 GCACAGCAAGTCAACCCGAACACCAITTCACAGTTTGTAAACAGAGTGGATGAAGATAGC 1129  
Db 2425 CCACGGCAAGTCTCCCGAGGACCACCTGTATGTGACTGTGAGTTGGAGAAACAGTAGC 2483

RESULT 4

US-08-828-008-1  
; Sequence 1, Application US/08828008  
; Patent No. 5939316  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Friemer, Nelson, B.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC  
; TITLE OF INVENTION: DISORDERS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/828,008  
; FILING DATE: 27-MAR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-071  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1144 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; NAME/KEY: Coding Sequence

; LOCATION: 1...282  
; OTHER INFORMATION:  
; NAME/KEY: Fsh 22  
; LOCATION: 1...1144  
; OTHER INFORMATION:  
; US-08-828-008-1  
  
Query Match 19.9%; Score 256.4; DB 2; Length 1144;  
Best Local Similarity 67.2%; Pred. No. 5.8e-71;  
Matches 437; Conservative 0; Mismatches 206; Indels 7; Gaps 5;  
  
Qy 480 GAAGCTGGAATGGAATGTGAGCGGTGTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCA 539  
Db 63 GAAGTCCGATGCTGATGGCAGTAGCTTTGAGGAGGCGTTGAGCAGAGCTGTCAACCGTCCA 122  
Qy 540 TAAAGAAAGAGCGCGCAGTCCAACTGGNAATCCTCAGCATCTCTGATTTCTCAAGCCCA 599  
Db 123 CAAGAGGAAGGCGCGCAGTCCAACTGGAAACCGGAGCA-CTTCTCATCTCGAGGCCCA 181  
Qy 600 GTTTCCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGTCTGTGATCTGGGCC 659  
Db 182 GTTCGCTCGAGCTTGGGGAGACACAGAGGGCAAGTACATCATGTTCGACTTTGGGCC 241  
Qy 660 ACAAGAGCGTATGCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTG 719  
Db 242 ACAGGAGAGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACCATCAGCCACTG 301  
Qy 720 GCTGGCCAACGTCNAAGTACCAGCTTAGGAAACGGCGGGACAAAATTTCTGAAAACAT 779  
Db 302 GCTGGCCAATGTGAAGTACCAGTTGAGGAGGACAGGGGGAAAGAAATCTTAAAGAACCT 361  
Qy 780 GGACAAAGGCCACCCCATCTTTTATTGCACTGTGCTGCCAGTTCAGAACCCCTTC 839  
Db 362 GGACACAGGGCATCTGTTTCTTTTGGCAACATGTGCTCTCAGTTGAGAACTGCTTC 421  
Qy 840 TACCTACATCAGTCACTTAGAATCTCACCTGGTTTCCAAATGAAGGACATGACCCGCTT 899  
Db 422 TACATACATAAGTCACTTTGGAGACACACTTGGCTTCAGCTTGAAGATCTCTCCAAGCT 481  
Qy 900 GTCAGTGGACAGCAAAAGCAAGGTGAGCAAGAGATCTCCCGGTATCTCGGCTCAGAG 959  
Db 482 GCCACT---CAATCAGATTCAAGAACAGACAGAAATGTTTCAAAGTCTCTCACCACAAAC 538  
Qy 960 GTCTCCAGAAACAATAGTCTGCCGAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTG 1019  
Db 539 TCTGGGCCCACTGGGGGCCACCGAGGAAGACTTGGCTCCACATTCATGTAAGTCTG 598  
Qy 1020 CTGTGGACATTTGTGAGCAAAACATGCGGTAAACCTCCACCTTAAGCAAAACGACAGCAA 1079  
Db 599 CAACGGACTTTTG-GAGCAAGCAGCAGTCAAA-TGCACCTTAGTAAGACCCACGCGC-A 655  
Qy 1080 GTCACCCGAACACCATTCACAGTTTGTAAACAGCGTGGATGAAGATAGC 1129  
Db 656 GTCTCCCGAGGACCACCTGTATGTGACTGTATGTGAGTTGGAGAAACAGTAGC 705

RESULT 5

US-08-828-008-3  
; Sequence 3, Application US/08828008  
; Patent No. 5939316  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; APPLICANT: Friemer, Nelson, B.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC  
; TITLE OF INVENTION: DISORDERS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA



ZIP: 10036/2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: Fast-SEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/828,008  
FILING DATE: 27-MAR-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-071  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 282 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
FEATURE:  
NAME/KEY: Other  
LOCATION: 1...282  
OTHER INFORMATION: Open reading frame to  
OTHER INFORMATION: Fsh 22  
US-08-828-008-3

Query Match 8.8%; Score 113; DB 2; Length 282;  
Best Local Similarity 72.4%; Pred. No. 1.4e-25;  
Matches 160; Conservative 0; Mismatches 60; Indels 1; Gaps 1;  
Qy 480 GAAGTCGAAATGGATGTCCAGCGCTTTTGGAGATCTCCAGTCAAGTCAACTTTGCA 539  
Db 63 GAAGTCGATGCTGATGCGAGTAGTTTGGAGGCGGTGGAGAGCTGTCAACGGTCCA 122  
Qy 540 TAAAGAAAGCGCGAGTCCAACTGGAATCTCAGCATCTTCTGATTTACAGGCCCA 599  
Db 123 CAAGAGGAAGCGCGGAGTCCAACTGGAACCCGAGCA-CTTCTCATCTCGAGGCCCA 181  
Qy 600 GTTTCCTCGAGCTTCTCCAGACATCAGAGGGGAAATACCTGCTGTGATCTGGGCC 659  
Db 182 GTTCGCTCGAGCTTGGCGGGAGACCACAGAGGGCAAGTACATCATGTCGGACTTGGGCC 241  
Qy 660 ACAAGAGGTATGCAAACTCTAAGTTTACGGGACTCTCAA 700  
Db 242 ACAGGAGGGTGCACTCTGAAGTTTACTGGGCTCTCCA 282

RESULT 6  
US-09-023-655-773  
Sequence 773, Application US/09023655  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Seilhaer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA

ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 773:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 258 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PGANOT01  
CLONE: 626570  
US-09-023-655-773

Query Match 3.2%; Score 41.6; DB 4; Length 258;  
Best Local Similarity 88.0%; Pred. No. 0.0086;  
Matches 44; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
Qy 1 GAGGGCAGCGAGAGGAGAAACCCAGCCCTCGAGGCCACCATCTGCTCT 50  
Db 207 GAGGGCAGCGAGAGGAGAAACCCAGCCCTCGAGGCCACCATCTTNTT 256

RESULT 7  
US-09-949-016-15945  
Sequence 15945, Application US/09949016  
Patent No. 6812339  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig et al.  
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: CL001307  
CURRENT APPLICATION NUMBER: US/09/949,016  
CURRENT FILING DATE: 2000-04-14  
PRIOR APPLICATION NUMBER: 60/241,755  
PRIOR FILING DATE: 2000-10-20  
PRIOR APPLICATION NUMBER: 60/237,768  
PRIOR FILING DATE: 2000-10-03  
PRIOR APPLICATION NUMBER: 60/231,498  
PRIOR FILING DATE: 2000-09-08  
NUMBER OF SEQ ID NOS: 207012  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15945  
LENGTH: 31623  
TYPE: DNA  
ORGANISM: Human  
US-09-949-016-15945

Query Match 3.1%; Score 40; DB 4; Length 31623;  
Best Local Similarity 54.9%; Pred. No. 0.66;  
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
Qy 362 CACAATCTGTATGTCCCACTCAGAGAGCGCTCTCTGACATCCCGACATGTC 421  
Db 28477 CACACCTTAGCATGGCTTCCCGAGGGGCCAGGTTCTGGCTCATGTCTCCCATTCGGCC 28536



```
; ORGANISM: Human
US-09-949-016-115812

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCCTCTGACACCTGGCTCTGAGAAG 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 ACTGCCAAAAA AAAAAAAAAA 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AATACAAAAA AAAAAAAAAA 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-09-949-016-115905/c
; Sequence 115905, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 115905
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-115905

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCCTCTGACACCTGGCTCTGAGAAG 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 ACTGCCAAAAA AAAAAAAAAA 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AATACAAAAA AAAAAAAAAA 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
US-09-949-016-115998/c
; Sequence 115998, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 115998
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-115998

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCCTCTGACACCTGGCTCTGAGAAG 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 ACTGCCAAAAA AAAAAAAAAA 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AATACAAAAA AAAAAAAAAA 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
US-09-949-016-116091/c
; Sequence 116091, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-08
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 116091
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-116091

Query Match          2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28;
Matches 55; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCCTCTGACACCTGGCTCTGAGAAG 1259
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 ACCATCCTGGCTAAACCGTGAACCCCTGGCTAAACCGTGAACCCCTGTCTCTACTAAA 325
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1260 ACTGCCAAAAA AAAAAAAAAA 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AATACAAAAA AAAAAAAAAA 301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 15
US-09-949-016-116184/c
; Sequence 116184, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/237,768
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 116184
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-116184

Query Match      2.9%; Score 37.6; DB 4; Length 601;
Best Local Similarity 65.5%; Pred. No. 0.28; Indels 0; Gaps 0;
Matches 55; Conservative 0; Mismatches 29;

QY 1200 ACCATCCTGCTTCTGACATTGAACCTCATTGAACCTCCTCTGACACCCCTGGCTCTGAGAAG 1259
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 384 ACCATCCTGGCTAAACGGTGAACCCCTGGCTAAACGGTGAACCCCTGCTCTACTATAA 325
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1260 ACTGCCAAAAA AAAAAAAAAA 1283
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AATACAAAAA AAAAAAAAAA 301
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 5, 2005, 07:03:17  
Job time : 293.689 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2005, 03:59:12 ; Search time 1770.09 Seconds  
(without alignments)  
4449.254 Million cell updates/sec

Title: US-08-731-499-5

Perfect score: 1288

Sequence: 1 GAGGCGCGGAGAGGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters: 11308400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*
- 17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*
- 18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*
- 19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*
- 20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*
- 21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*
- 22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1288	100.0	1288	8	US-08-731-499-5
2	1265	98.2	3322	17	US-10-104-047-373
3	773	60.0	2449	9	US-09-764-864-628
4	770.4	59.8	799	9	US-09-764-864-193
5	656.4	51.0	2012	9	US-09-864-761-33203
6	521	40.5	1964	9	US-09-864-761-16736
7	390.4	30.3	3604	10	US-09-764-891-9420
8	390.4	30.3	3604	17	US-10-091-414-288
9	390.4	30.3	3907	9	US-09-823-245A-409
10	272	21.1	3692	18	US-10-723-860-2953
11	272	21.1	5065	18	US-10-723-860-5900

12	272	21.1	5065	18	US-10-723-860-6999	Sequence 6999, Ap
13	272	21.1	5065	18	US-10-723-860-7288	Sequence 7288, Ap
14	256.4	19.9	1144	9	US-09-739-707-1	Sequence 1, Appli
15	214.2	16.6	4839	17	US-10-062-674-1795	Sequence 1795, Ap
16	157	12.2	346	10	US-09-960-706-273	Sequence 273, App
17	157	12.2	346	10	US-09-873-319-165	Sequence 165, App
18	143.4	11.1	2044	9	US-09-764-864-137	Sequence 137, App
19	143	11.1	852	9	US-09-764-864-576	Sequence 576, App
20	142.8	11.1	443	9	US-09-796-692-5430	Sequence 5430, Ap
21	142.8	11.1	443	14	US-10-040-862-5430	Sequence 5430, Ap
22	142.8	11.1	443	17	US-10-057-475B-5430	Sequence 5430, Ap
23	142.8	11.1	443	17	US-10-154-884B-5430	Sequence 5430, Ap
24	142.8	11.1	443	18	US-10-764-324-5430	Sequence 5430, Ap
25	133.8	10.4	350	17	US-10-062-674-56	Sequence 56, Appl
26	113	8.8	282	9	US-09-739-707-3	Sequence 3, Appli
27	105	8.2	3370	17	US-10-104-047-338	Sequence 338, App
28	97.2	7.5	450	10	US-09-918-995-164	Sequence 164, App
29	65.2	5.1	537	10	US-09-918-995-19673	Sequence 19673, A
30	51.6	4.0	640	14	US-10-198-846-13601	Sequence 13601, A
31	51.6	4.0	904	14	US-10-198-846-7085	Sequence 7085, Ap
32	42.6	3.3	497	9	US-09-783-590-5119	Sequence 5119, Ap
33	41.6	3.2	258	17	US-10-641-643-773	Sequence 773, App
34	40.2	3.1	653	14	US-10-184-644-402	Sequence 402, App
35	40.2	3.1	653	14	US-10-184-634-402	Sequence 402, App
36	40	3.1	515	16	US-10-029-386-6506	Sequence 6506, Ap
37	39.4	3.1	1796	18	US-10-425-115-72371	Sequence 72371, A
38	39.2	3.0	289730	18	US-10-719-993-6780	Sequence 6780, Ap
39	38.6	3.0	703	13	US-10-027-632-174146	Sequence 174146,
40	38.6	3.0	703	17	US-10-027-632-174146	Sequence 174146,
41	38.2	3.0	2886	15	US-10-156-761-5473	Sequence 5473, Ap
42	38.2	3.0	9025608	15	US-10-156-761-1	Sequence 1, Appli
43	38	3.0	594	14	US-10-123-155-10	Sequence 10, Appl
44	38	3.0	594	15	US-10-146-731-10	Sequence 10, Appl
45	38	3.0	594	15	US-10-140-472-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-08-731-499-5  
; Sequence 5, Application US/08731499  
; Publication No. US20030148270A1  
; GENERAL INFORMATION:  
; APPLICANT: GRAY, Joe W.  
; APPLICANT: COLLINS, Colin  
; APPLICANT: HWANG, Soo-In  
; APPLICANT: GODFREY, Tony  
; APPLICANT: KOWBEL, David  
; APPLICANT: KOMMENS, Johanna  
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR  
; TITLE OF INVENTION: USES  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/731,499  
; FILING DATE: 16-OCT-1996  
; CLASSIFICATION: 435  
; PRIORITY APPLICATION DATA:  
; APPLICATION NUMBER: US 08/680,395  
; FILING DATE: 15-JUL-1996  
; ATTORNEY/AGENT INFORMATION:

```
/ NAME: Hunter, Tom
/ REGISTRATION NUMBER: 38,498
/ REFERENCE/DOCKET NUMBER: 23070-068910
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1288 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: - 1.1288
/ LOCATION: 1..1288
/ OTHER INFORMATION: /note= "cDNA clone 41.1 with homology
/ OTHER INFORMATION: to homeobox T shirt gene from
/ OTHER INFORMATION: Drosophila"
US-08-731-499-5

Query Match          100.0%; Score 1288; DB 8; Length 1288;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60
DB 1 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60

QY 61 TGGCCCTCGCCAAACCAACGAGCCCGGCGCTGCCATGATCAACCACTCAGCGCCCTGCAG 120
DB 61 TGGCCCTCGCCAAACCAACGAGCCCGGCGCTGCCATGATCAACCACTCAGCGCCCTGCAG 120

QY 121 TCGTCCTGAACATCACTTGGGCAAGCAACGAGCCCTGGGCTCACTTCTCTGCTCC 180
DB 121 TCGTCCTGAACATCACTTGGGCAAGCAACGAGCCCTGGGCTCACTTCTCTGCTCC 180

QY 181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAAGTCGAATCTCAATGTATGGACAAG 240
DB 181 AGCCCAAGTTCAAGCACAATTTCCATGTTCCACAAGTCGAATCTCAATGTATGGACAAG 240

QY 241 CCGGTCTTGAGTCTGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGGAG 300
DB 241 CCGGTCTTGAGTCTGCTCCCAAGGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGGAG 300

QY 301 ACAGCGATCAGCCCACTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCGCAA 360
DB 301 ACAGCGATCAGCCCACTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCGCAA 360

QY 361 GCACAATCTTGATGTCCCACTCAGAAGCAGCTCTGTCTGACATCGCCGACATGGTC 420
DB 361 GCACAATCTTGATGTCCCACTCAGAAGCAGCTCTGTCTGACATCGCCGACATGGTC 420

QY 421 AAGTCTCCCCAAGCAACCAACCCAGCCAGCTCTCTCTCCAGGCTCCCCCCATG 480
DB 421 AAGTCTCCCCAAGCAACCAACCCAGCCAGCTCTCTCTCCAGGCTCCCCCCATG 480

QY 481 AAGCTGGAATGATGTGAGGCGCTTGGAGATGTCTCCAGTCAAGTCTCAACTTGCAT 540
DB 481 AAGCTGGAATGATGTGAGGCGCTTGGAGATGTCTCCAGTCAAGTCTCAACTTGCAT 540

QY 541 AAAAGAAAAGCGCGAGTCCAACTCGAATCTTCAGCATCTTCTGATTTCTACAAGCCAG 600
DB 541 AAAAGAAAAGCGCGAGTCCAACTCGAATCTTCAGCATCTTCTGATTTCTACAAGCCAG 600

QY 601 TTGTCCTCGAGCCTCTTCCAGACATCAGAGGCAAAATACCTGTCTGTATCTGGGCCCA 660
DB 601 TTGTCCTCGAGCCTCTTCCAGACATCAGAGGCAAAATACCTGTCTGTATCTGGGCCCA 660

QY 661 CAAGCGGTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTACTCAGTCACTGG 720
DB 661 CAAGCGGTATGCAATCTTAAGTTTACGGGACTCTCAATGACCACTACTCAGTCACTGG 720
```

```
QY 721 CTGGCCAAACGTCAAGTACAGCTTAGGAAAACGGGGCGGACAAAATTTCTGAAAAACATG 780
DB 721 CTGGCCAAACGTCAAGTACAGCTTAGGAAAACGGGGCGGACAAAATTTCTGAAAAACATG 780

QY 781 GACAAAGGCCACCCCATCTTTTATTCAGTACTGTCCTCCAGTTTCAGAACCCCTTCT 840
DB 781 GACAAAGGCCACCCCATCTTTTATTCAGTACTGTCCTCCAGTTTCAGAACCCCTTCT 840

QY 841 ACCTACATCAGTCACCTTAGAATCTCACCTGGGTTCCTCAATGAAGGACATGACCCGCTTG 900
DB 841 ACCTACATCAGTCACCTTAGAATCTCACCTGGGTTCCTCAATGAAGGACATGACCCGCTTG 900

QY 901 TCAGTGGACCAAGCAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCGTCGGCTCAGAGG 960
DB 901 TCAGTGGACCAAGCAAGCAAGGTGGAGCAAGAGATCTCCCGGTATCGTCGGCTCAGAGG 960

QY 961 TCTCCAGAAAACAATAGCTGCCGAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020
DB 961 TCTCCAGAAAACAATAGCTGCCGAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGC 1020

QY 1021 TGTTCGGACATTTGTGAGCAAAACATCGGTAAAACTCCACCTAAGCAAAAACGACAGCAAG 1080
DB 1021 TGTTCGGACATTTGTGAGCAAAACATCGGTAAAACTCCACCTAAGCAAAAACGACAGCAAG 1080

QY 1081 TCACCCGAAACACCAATTCACAGTTTGTAAACAGACGTGGATGAAAGATAGCTCTGCAGGACG 1140
DB 1081 TCACCCGAAACACCAATTCACAGTTTGTAAACAGACGTGGATGAAAGATAGCTCTGCAGGACG 1140

QY 1141 AATGCTTAGTTTCACTTCCAGCTTGGATGCCCTCACACTGAACCCCTTCTCGTTGCA 1200
DB 1141 AATGCTTAGTTTCACTTCCAGCTTGGATGCCCTCACACTGAACCCCTTCTCGTTGCA 1200

QY 1201 CCATCTCTGCTTGACATTTGAACTCAATTCCTCTGACACACCTGCTCTGAGAAGA 1260
DB 1201 CCATCTCTGCTTGACATTTGAACTCAATTCCTCTGACACACCTGCTCTGAGAAGA 1260

QY 1261 CTGCCAAAAAATAAAAAAAAAAAAAATTC 1288
DB 1261 CTGCCAAAAAATAAAAAAAAAAAAAATTC 1288

RESULT 2
US-10-104-047-373
; Sequence 373, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 373
; LENGTH: 3322
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-373

Query Match          98.2%; Score 1265; DB 17; Length 3322;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 60
DB 2058 GAGGGCAGCGAGAGAGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGG 2117

QY 61 TCGGCCCTCGCCAAACCAACGAGCCCGGCTTGCATGATCAACCACTCAGCGCCCTGCAG 120
DB 2118 TCGGCCCTCGCCAAACCAACGAGCCCGGCTTGCATGATCAACCACTCAGCGCCCTGCAG 2177
```

Qy	121	TCGGTCTCTGAA	CAATCACTTGGGCA	AAGCCACGAGCCCTTGGCGTCA	CCCTTCTGCTCC	180
Db	2178	TCGGTCTCTGAA	CAATCACTTGGGCA	AAGCCACGAGCCCTTGGCGTCA	CCCTTCTGCTCC	2237
Qy	181	AGCCCAAGTTC	CAAGCACAATTTCCATGTTCCACAAGTTCGAATCTCAATGTCA	TGATCGATCGACAAG	240	
Db	2238	AGCCCAAGTTC	CAAGCACAATTTCCATGTTCCACAAGTTCGAATCTCAATGTCA	TGATCGACAAG	2297	
Qy	241	CCGGTCTTTAGT	CTCTGCTCCACAAGGTTCAGCCAGCGTGTCCAGCGCGTACCTCTGTTTGAG	300		
Db	2298	CCGGTCTTTAGT	CTCTGCTCCACAAGGTTCAGCCAGCGTGTCCAGCGCGTACCTCTGTTTGAG	2357		
Qy	301	AACAGCGATCA	GCCCATTTGACTGACCAAGTTCAAAAAGCAAGAACGCGAGTCTCTCGAA	360		
Db	2358	AACAGCGATCA	GCCCATTTGACTGACCAAGTTCAAAAAGCAAGAACGCGAGTCTCTCGAA	2417		
Qy	361	GCACAACTTTGAT	GTGTCCTCCACCTTCAGAACGACGCTGTGTCTGACATCGCGGACATGGTC	420		
Db	2418	GCACAACTTTGAT	GTGTCCTCCACCTTCAGAACGACGCTGTGTCTGACATCGCGGACATGGTC	2477		
Qy	421	AAAGTCTCTCCC	AAAAAGCCACACCCCAAGGCAGCTCTCTCCACAGGGTCCCCCCCAGT	480		
Db	2478	AAAGTCTCTCCC	AAAAAGCCACACCCCAAGGCAGCTCTCTCCAGGGTCCCCCCCAGT	2537		
Qy	481	AAGCTGGAAT	TGGATGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT	540		
Db	2538	AAGCTGGAAT	TGGATGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTTGCAT	2597		
Qy	541	AAAAAGAAAGC	CGGCGAGTCCAACTGGNAATCCTCAGCATCTTCTGATCTTACAGGCCAG	600		
Db	2598	AAAAAGAAAGC	CGGCGAGTCCAACTGGNAATCCTCAGCATCTTCTGATCTTACAGGCCAG	2657		
Qy	601	TTTGCTCTCAG	CCCTCTCCAGACATCAGAGGCAAAATACCTGCTGTGTGATCTGGGCCCA	660		
Db	2658	TTTGCTCTCAG	CCCTCTCCAGACATCAGAGGCAAAATACCTGCTGTGTGATCTGGGCCCA	2717		
Qy	661	CAAGAGCGTAT	GCAAACTCTTAAGTTTACGCGGACTCTCAATGACCACTATCAGTCACTGG	720		
Db	2718	CAAGAGCGTAT	GCAAACTCTTAAGTTTACGCGGACTCTCAATGACCACTATCAGTCACTGG	2777		
Qy	721	CTGGCCAAAGT	CAAGTACCAGCTTAGGNAAGCGGCGGACAAAATTTCTGAAAAAATG	780		
Db	2778	CTGGCCAAAGT	CAAGTACCAGCTTAGGNAAGCGGCGGACAAAATTTCTGAAAAAATG	2837		
Qy	781	GACAAAGCCAC	CCCCACTCTTTTATTTCAGTGCATGTGCTCCCAAGTTTCAGAAACCCCTTCT	840		
Db	2838	GACAAAGCCAC	CCCCACTCTTTTATTTCAGTGCATGTGCTCCCAAGTTTCAGAAACCCCTTCT	2897		
Qy	841	ACCTACATCAG	TCACTTAGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGCTTG	900		
Db	2898	ACCTACATCAG	TCACTTAGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGCTTG	2957		
Qy	901	TCAGTGGACAG	AAAGCAAGGTGGAGCAAGATCTCTCCCGGGTATCGTGGGCTCAGAGG	960		
Db	2958	TCAGTGGACAG	AAAGCAAGGTGGAGCAAGATCTCTCCCGGGTATCGTGGGCTCAGAGG	3017		
Qy	961	TCTCCAGAAA	CAATAGCTGCGGACAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTC	1020		
Db	3018	TCTCCAGAAA	CAATAGCTGCGGACAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTC	3077		
Qy	1021	TGTCGGACAT	TTGTGAGCAAACTGCGGTAAAACTCCACCTTAAGCAAAACGACAGCAAG	1080		
Db	3078	TGTCGGACAT	TTGTGAGCAAACTGCGGTAAAACTCCACCTTAAGCAAAACGACAGCAAG	3137		
Qy	1081	TCACCCGAAC	ACCATTCACAGTTGTATACAGAGTGGATGAAGAAATAGTCTCTGAGGACG	1140		
Db	3138	TCACCCGAAC	ACCATTCACAGTTGTGTATACAGAGTGGATGAAGAAATAGTCTCTGAGGACG	3197		
Qy	1141	AATGCTTAGT	TTTCCACTTTCCAGCTCGATCCCTCCTACACTGAAACCCCTTCTCGTTGCA	1200		
Db	3198	AATGCTTAGT	TTTCCACTTTCCAGCTCGATCCCTCCTACACTGAAACCCCTTCTCGTTGCA	3257		
Qy	1201	CCATCCTCTG	CTGACATTTGAATCACTTGAATCTCCTCTGACACCCCTGGCTCTCAGAGA	1260		

```

Db      3258  CCATCTCGCTTCTGACATTGAAGCTCCTGTGACACCCCTGGCTCTGAGAAGA   3317
|||||
Qy      1261  CTGCC 1265
|||||
Db      3318  CTGCC 3322

RESULT 3
US-09-764-864-628
; Sequence 628, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 628
; LENGTH: 2449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (13)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2437)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-864-628
```

Query Match	60.0%;	Score 773;	DB 9;	Length 2449;
Best Local Similarity	100.0%;	Pred. No. 4.3e-231;		
Matches 773;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	514	GTCTCCAGTGAAGTCTCAACTTTTGCATAAAGAAAGCCGGCAGTCCCACTGGAATCTCT	573	
Db	95	GTCTCCAGTGAAGTCTCAACTTTTGCATAAAGAAAGCCGGCAGTCCCACTGGAATCTCT	154	
Qy	574	CAGCATCTTCTGATTCTTACAAGCCCAAGTTGCGCTCGAGCCCTCTTCCAGACATCAGAGGCG	633	
Db	155	CAGCATCTTCTGATTCTTACAAGCCCAAGTTGCGCTCGAGCCCTCTTCCAGACATCAGAGGCG	214	
Qy	634	AAATACCTGTCTGTATCTCTGGGCCCAAGAGGGTATGCAAAATCTCTTAAGTTTACGGGA	693	
Db	215	AAATACCTGTGTCTGATCTGGGCCCAAGAGGGTATGCAAAATCTCTTAAGTTTACGGGA	274	
Qy	694	CTCTCAATGACCACTATCAGTCACTGGCTGGCCAAACGTCAGTACCAGCTTTAGGAAACG	753	
Db	275	CTCTCAATGACCACTATCAGTCACTGGCTGGCCAAACGTCAGTACCAGCTTTAGGAAACG	334	
Qy	754	GGCGGGACAAAATTTCTGAAAAACATCGGACAAAGGCCACCCCATCTTTTATTGCAGTGAC	813	
Db	335	GGCGGGACAAAATTTCTGAAAAACATCGGACAAAGGCCACCCCATCTTTTATTGCAGTGAC	394	
Qy	814	TGTGCTCCCAAGTTCAGAAACCCCTTCTACCTPACATCAGTCACTTAGAATCTCACTGGGT	873	
Db	395	TGTGCTCCCAAGTTCAGAAACCCCTTCTACCTPACATCAGTCACTTAGAATCTCACTGGGT	454	
Qy	874	TTCCAAATGAGGACATGACCGCTTCTCAGTGGACCAAGCAAGGTCGGAGCAAGAG	933	
Db	455	TTCCAAATGAGGACATGACCGCTTCTCAGTGGACCAAGCAAGGTCGGAGCAAGAG	514	
Qy	934	ATCTCCCGGGTATCGTCCGGCTCAGAGGTCTCCAGAAAAAATAGTCTGCCGAGAGGACACA	993	
Db	515	ATCTCCCGGGTATCGTCCGGCTCAGAGGTCTCCAGAAAAAATAGTCTGCCGAGAGGACACA	574	
Qy	994	GACTCTAAATTCAGAGTGAAGTTGTGTGCGGACATTTGTGAGCAAAACATGCGGTAAAA	1053	
Db	575	GACTCTAAATTCAGAGTGAAGTTGTGTGCGGACATTTGTGAGCAAAACATGCGGTAAAA	634	

```

QY 1054 CTCACCTAAGCAAAACGACAGCAAGTCACCCGAACACCAATTCACAGTTTGTAAACAGAC 1113
Db 635 CTCACCTAAGCAAAACGACAGCAAGTCACCCGAACACCAATTCACAGTTTGTAAACAGAC 694
QY 1114 GTGGATGAAGAAATAGCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 1173
Db 695 GTGGATGAAGAAATAGCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 754
QY 1174 CCTCACCTGAACCTCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 1233
Db 755 CCTCACCTGAACCTCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 814
QY 1234 CCTCTCGACACCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 1286
Db 815 CCTCTCGACACCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 867

RESULT 4
US-09-764-864-193
; Sequence 193, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 192
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 193
; LENGTH: 799
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-193

Query Match 59.8%; Score 770.4; DB 9; Length 799;
Best Local Similarity 99.9%; Pred. No. 1.7e-230;
Matches 771; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 514 GTCTCAGTGAAGTCTCAACTTTTGCATATAAAGAAAGCGCGAGTCCAACTGGAAATCT 573
Db 12 GTCTCAGTGAAGTCTCAACTTTTGCATATAAAGAAAGCGCGAGTCCAACTGGAAATCT 71
QY 574 CAGCATCTCTGATCTACAGCCAGTTTGCCTCGAGCCTCTCCAGACATCAGAGGCG 633
Db 72 CAGCATCTCTGATCTACAGCCAGTTTGCCTCGAGCCTCTCCAGACATCAGAGGCG 131
QY 634 AAATACCTCTGCTGATCTGGGCCCCACAAAGCGGTATGCAATCTCTAAAGTTTACGGGA 693
Db 132 AAATACCTCTGCTGATCTGGGCCCCACAAAGCGGTATGCAATCTCTAAAGTTTACGGGA 191
QY 694 CTCTCAATGACCAATATAGTCACTGGTGGGCAAGTCAAGTACAGCTTAAAGAAAGCG 753
Db 192 CTCTCAATGACCAATATAGTCACTGGTGGGCAAGTCAAGTACAGCTTAAAGAAAGCG 251
QY 754 GCGGGGACAAATTTCTGAABACATGGGCAAGGCCACCCATCTTTTATTCAGTGC 813
Db 252 GCGGGGACAAATTTCTGAABACATGGGCAAGGCCACCCATCTTTTATTCAGTGC 311
QY 814 TGTGCTCCAGTTCAGAAACCCCTCTTCACTACATCAGTCACTTAGAATCTCACCTGGGT 873
Db 312 TGTGCTCCAGTTCAGAAACCCCTCTTCACTACATCAGTCACTTAGAATCTCACCTGGGT 371
QY 874 TTCCAAATGAAGACATGACCCGCTTTGTGAGTGGACCAAGCAAGAGTGGAGCAAGAG 933
Db 372 TTCCAAATGAAGACATGACCCGCTTTGTGAGTGGACCAAGCAAGAGTGGAGCAAGAG 431
QY 934 ATCTCCCGGTATCTCGGCTCAGAGTCTCCAGAAACAATAGCTGCCGAGAGACACA 993
Db 432 ATCTCCCGGTATCTCGGCTCAGAGTCTCCAGAAACAATAGCTGCCGAGAGACACA 491

```

```

QY 994 GACTCTAAATTCAGTGTAAAGTTGTGCTGTGGACATTTGTGAGCAAAACATCGGTTAAAA 1053
Db 492 GACTCTAAATTCAGTGTAAAGTTGTGCTGTGGACATTTGTGAGCAAAACATCGGTTAAAA 551
QY 1054 CTCACCTAAGCAAAACGACAGCAAGTCACCCGAACACCAATTCACAGTTTGTAAACAGAC 1113
Db 552 CTCACCTAAGCAAAACGACAGCAAGTCACCCGAACACCAATTCACAGTTTGTAAACAGAC 611
QY 1114 GTGGATGAAGAAATAGCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 1173
Db 612 GTGGATGAAGAAATAGCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 671
QY 1174 CCTCACCTGAACCTCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 1233
Db 672 CCTCACCTGAACCTCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 731
QY 1234 CCTCTCGACACCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 1285
Db 732 CCTCTCGACACCTCTGAGGAGCAATGCTTTAGTTTCCATTTCCAGCTCGATCC 783

RESULT 5
US-09-864-761-33203
; Sequence 33203, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```



```

; SEQ ID NO 33203
; LENGTH: 2012
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004505.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: NT HIT: AF207880.1, EVALU0.0.00e+00
; OTHER INFORMATION: EST HUMAN HIT: H12950.1, EVALU0.0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P22265, EVALU0.0.00e+07
US-09-864-761-33203

Query Match          51.0%; Score 656.4; DB 9; Length 2012;
Best Local Similarity 99.8%; Pred. No. 1.4e-194;
Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1  GAGGGCAGCAGAGAGAGAAACCCAGCCCTGGAGCCCAATCTGCTCTGAGCAATGGG 60
DB      1347 GAGGGCAGCAGAGAGAGAAACCCAGCCCTGGAGCCCAATCTGCTCTGAGCAATGGG 1406

QY      61  TGGCGCCTCGCCAAACCAACGCGCCGCGCCCTGCCATGCAATCAACCCACTCAGCGCCCTGCAG 120
DB      1407 TGGCGCCTCGCCAAACCAACGCGCCGCGCCCTGCCATGCAATCAACCCACTCAGCGCCCTGCAG 1466

QY      121 TCGGTCTTGAAACAAATCACTTGGGCAAGACGAGCGCCCTTGGCGTCACCTTCTGTGCTCC 180
DB      1467 TCGGTCTTGAAACAAATCACTTGGGCAAGACGAGCGCCCTTGGCGTCACCTTCTGTGCTCC 1526

QY      181 AGCCCAAGTTCACAGCAAAATTCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAG 240
DB      1527 AGCCCAAGTTCACAGCAAAATTCATGTTCCACAAGTCGAATCTCAATGTCATGGACAAG 1586

QY      241 CCGGTCTTGAGTCTCTGCTCTCACAAAGTCAAGCGTGTCCAGGCGCTACCTGTTTGAG 300
DB      1587 CCGGTCTTGAGTCTCTGCTCTCACAAAGTCAAGCGTGTCCAGGCGCTACCTGTTTGAG 1646

QY      301 AACAGCGATCAGGCCATGTGACCTGACCAAGTCCAAAAGCAAGAACGCGAGTCTCTCGCAA 360
DB      1647 AACAGCGATCAGGCCATGTGACCTGACCAAGTCCAAAAGCAAGAACGCGAGTCTCTCGCAA 1706

QY      361 GCACAAATCTTGATGTCCTCCACCTCAGAACGACGCTCTGTGTGACATCGCGGACATGGTC 420
DB      1707 GCACAAATCTTGATGTCCTCCACCTCAGAACGACGCTCTGTGTGACATCGCGGACATGGTC 1766

QY      421 AAAGTCTCTCCCAAAGCCACACCCCAAGCAGCCCTCTCTCTCCAGGGTCCCCCCCCATG 480
DB      1767 AAAGTCTCTCCCAAAGCCACACCCCAAGCAGCCCTCTCTCTCCAGGGTCCCCCCCCATG 1826

QY      481 AAGCTGGAATTGGATGTGAGCGGCTTTTGAGATGTCTCCAGTGAAGTCTCAACTTTGCAT 540
DB      1827 AAGCTGGAATTGGATGTGAGCGGCTTTTGAGATGTCTCCAGTGAAGTCTCAACTTTGCAT 1886

QY      541 AAAAAGAAAGCCCGGAGTCCAACCTGAATCTCAGCATCTTCTGATTTACAGGCCAG 600
DB      1887 AAAAAGAAAGCCCGGAGTCCAACCTGAATCTCAGCATCTTCTGATTTACAGGCCAG 1946

QY      601 TTTGSCCTCGAGCCCTTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCC 658
DB      1947 TTTGSCCTCGAGCCCTTTCCAGACATCAGAGGGCAAAATACCTGCTGTCTGATCTGGGCC 2004

RESULT 6
US-09-864-761-16736
; Sequence 16736, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USE
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aesmica-x-1

```

```

; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16736
; LENGTH: 1964
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004505.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; US-09-864-761-16736

Query Match 40.5%; Score 521; DB 9; Length 1964;
Best Local Similarity 100.0%; Pred.No.4e-152;
Matches 521; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAGGCGAGCGAGGAAGGAGAAACCCACGCGCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG 60
Db 1444 GAGGCGAGCGAGGAAGGAGAAACCCACGCGCCCTGGAGCCCAACATCTGCTCTGAGCAATGGG 1503

Qy 61 TGGCGCCTCGCCAAACCAACGCGCGCGCCCTGGCCATGATCAACCCCACTCAGCGCCCTGCGAG 120
Db 1504 TGGCGCCTCGCCAAACCAACGCGCGCGCCCTGGCCATGATCAACCCCACTCAGCGCGCCCTGCGAG 1563

Qy 121 TCGGTCCTGAACAAATCAGTTGGCAAGCCACGAGCCCTTGGGCTCACCTTCCCTGCTGCC 180
Db 1564 TCGGTCCTGAACAAATCAGTTGGCAAGCCACGAGCCCTTGGGCTCACCTTCCCTGCTGCC 1623

Qy 181 AGCCCAAGTTCAGACCAATTTCCATGTTGCCAAGTTCGAATCTCAATGTCATGGACAAG 240
Db 1624 AGCCCAAGTTCAGACCAATTTCCATGTTGCCAAGTTCGAATCTCAATGTCATGGACAAG 1693

Qy 241 CCGGTTTGAGTCTCGCTCCCAACAGGTGAGCGGTGTCAGGCGCTACCTGTTTGGAG 300
Db 1684 CCGGTTTGAGTCTCGCTCCCAACAGGTGAGCGGTGTCAGGCGCTACCTGTTTGGAG 1743

```

QY 301 AACAGCGATCAGCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTCGAA 360  
 DB 1744 AACAGCGATCAGCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAGTCTCTCGAA 1803  
 QY 361 GCACAATCTTGATGTGTCCTCCACCTCAGAGCAGGCTCTGTGACATCGCCGACATGGTC 420  
 DB 1804 GCACAATCTTGATGTGTCCTCCACCTCAGAGCAGGCTCTGTGACATCGCCGACATGGTC 1863  
 QY 421 AAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTCCGCCCATG 480  
 DB 1864 AAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTCCGCCCATG 1923  
 QY 481 AAGCTGGAATGATGTCAGGCGCTTTGAGGATGTCCTCCAG 521  
 DB 1924 AAGCTGGAATGATGTCAGGCGCTTTGAGGATGTCCTCCAG 1964  
 RESULT 7  
 US-09-764-891-9420  
 ; Sequence 9420, Application US/09764891  
 ; Publication No. US20030077808A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC006  
 ; CURRENT APPLICATION NUMBER: US/09/764,891  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 10231  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9420  
 ; LENGTH: 3604  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (3443)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 US-09-764-891-9420  
 Query Match 30.3%; Score 390.4; DB 10; Length 3604;  
 Best Local Similarity 62.4%; Pred. No. 4.6e-111;  
 Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;  
 QY 52 AGCAATGGGTGGCGCTCGCCACCAACGCGCCGCTGCGCATGCAATCAACCACTCAGC 111  
 DB 1243 AACAACTGGGATCAATGACCACTCAGCGAGCCTTCTTCAATCAACCGCTGAGC 1302  
 QY 112 GCCCTGAGTCCGTCTCGAACAATCACTTTGGGCAAGCCACGAGCGCTTGGGCTCACCT 171  
 DB 1303 GCTTTGAGTCCATCAATCAACCACTGGGCAAGGTGCAAGCCGCTG----- 1353  
 QY 172 TCCTGCTCAGCCCAAGTTCAGCACAATTTCAATGTTCCAAAGTCGAATCTCAATGTC 231  
 DB 1354 -----AGTCCCTCGTGGACCGCTGGCGATGCTGTACAAATCAGCAACAGCATG 1404  
 QY 232 ATGGACAAGCGCTTTGAGTCTGCTGCTCCACAAAGTCAGCGCTGTCAGGCGCTAC 291  
 DB 1405 CTGGACAAGCGGTGTACCCCGCCACCCCTGTGGAAGCAGCGATGCCATCGACCGCTAC 1464  
 QY 292 CTGTTTGAACAAGCGATCAGCCATTGACCTGACCAAGTCCAAAGCAAGAAAGCCGAG 351  
 DB 1465 TATTATGAAAACAGCGACCGACCCATTGACTTAACCAAGTCCCAAGCAAGCGCTGGTG 1524  
 QY 352 TCCTCGCAAGCACAATCTTGATGTGTCCTCCCAAGCCACCTCAGAAAGCAGCTCTGTGACATCGCC 411  
 DB 1525 TCAGCGTGGCTGATTCGCTGGGATCACCTCTCGGGAGAGCGCACTCATGACATCTCC 1584  
 QY 412 GACATGGTCAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGTC 471  
 DB 1585 GACATGGTGAAGAACTCAGAGCCCGCTGACGCCCAAGTCTCTCCAGCCCTCCACAGTT 1644  
 QY 472 CCCCCCATGAAGCTGGAATGGATGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531

DB 1645 TCAGA---GAAGTCCGATGCTGATGCGAGCAGCTTTGAGGAGCGGTTGAGCAGAGCTGTCA 1701  
 QY 532 ACTTTGCAATAAAGAAAGGCGCGGAGTCCCACTGGAATCCTCAGCATCTTCTGATCTTA 591  
 DB 1702 CCGGTCCCAAGAGGAAGGCGCGGAGTCCCACTGGAACCCGAGCAGCTTCTCATCTCTG 1761  
 QY 592 CAAGCCAGTGTTCCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGTGTCTGAT 651  
 DB 1762 CAGGCCAGTTCGCTCGAGCTTGGGGAGACCAAGAGGCAAGTACATCATGTCTGGAC 1821  
 QY 652 CTGGGCCCAAGAGCGGTATGCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATC 711  
 DB 1822 TTGGGCCCGCAGAGAGGGGTGCATCTCGAAGTTTACTGGGCTCTCCATGACCCACATC 1881  
 QY 712 AGTCACTGGCTGGCCAAAGTCAAGTACAGCTTAGGAAACCGGGCGGGAATAATTTCTG 771  
 DB 1882 AGCCACTGGCTGGCCAAAGTGAAGTACCAAGTTGAGGAGGACAGGGGGAAACGAAATTTCTTA 1941  
 QY 772 AAAAAATGCAAAAGGCCACCCCATCTTTTATTCAGTGAAGTCTGTCCTCCAGTTTCTCA 831  
 DB 1942 AAGAACTGGACACAGGGGATCTCTGTTTCTTTTGAACGATGTGCTCTCAGTTTCTCA 2001  
 QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACCTGGGTTTCCAAATGAAGGACATG 891  
 DB 2002 ACTGCTTCTACATACATTAAGTCACTTTGGAGACACACTTGGGCTTCAAGGATCTC 2061  
 QY 892 ACCCGTGTGAGTGGACCAAGCAAGTGGAGCAAGATCTCCCGGTATCGTCG 951  
 DB 2062 TCCAAAGTCTCCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCGAAAGTCTCTCAC 2118  
 QY 952 GCTCAGAGGTCTCCAGAAACAATAGTCTGCCAAGAGGACACAGACTCTTAAATTTCAAGTGT 1011  
 DB 2119 AACAAACTCTGGGCGCACTGGGGGCCACCGAGGAGACTTGGGCTCCACATTTCAATGT 2178  
 QY 1012 AAGTTGTGCTGTCGACATTTCTGAGCAAAATGCGGTAAAACTCCACTAAGCAAAACG 1071  
 DB 2179 AAGCTCTGCAACCGGACTTTTTCGAGCAAGCAGCAGTCAAACTGACCTTAGTAAGACC 2238  
 QY 1072 CACAGCAAGTCCCGCAACCACTTACAGTTTGTAAACAGCGTGGATGAAGATAGC 1129  
 DB 2239 CAGGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 2296  
 RESULT 8  
 US-10-091-414-288  
 ; Sequence 288, Application US/10091414  
 ; Publication No. US20030224461A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: P1116C1  
 ; CURRENT APPLICATION NUMBER: US/10/091,414  
 ; CURRENT FILING DATE: 2002-03-07  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 392  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 288  
 ; LENGTH: 3604  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (3443)  
 ; OTHER INFORMATION: n equals a,t,g, or c  
 US-10-091-414-288  
 Query Match 30.3%; Score 390.4; DB 17; Length 3604;  
 Best Local Similarity 62.4%; Pred. No. 4.6e-111;  
 Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;  
 QY 52 AGCAATGGGTGGCGCTCGCCCAACCAAGCCCGGCTTCCCATGCAATCAACCACTCAGC 111

Db 1243 AACAACTGGGATCATATGACCACTACCGGAGCCTTCTTTCATCAACCGCTGAGC 1302  
Qy 112 GCCCTGAGTCGCTCTGAAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACT 171  
Db 1303 GCTTTGAGTCCATCATGAACACCCACCTGGGCAAGGTGTCCAAAGCCGCTG----- 1353  
Qy 172 TCCTGCTCCAGCCCAAGTTCAAGACAATTTCCATGTTCCCAAGTCGAATCTCAATGTC 231  
Db 1354 -----AGTCCCTCGTGACCGCTGGGATGCTGTACAAATCAGCAACAGCATG 1404  
Qy 232 ATGGACAAGCCGCTTCTGAGTCTGCTCCCAAGGTCAGCCAGCGGTGTCCAGGCGCTAC 291  
Db 1405 CTGGACAAGCCGCTGTACCCGCCACCCCTGTGAAGCAGCGGATGCCATCGACGCTAC 1464  
Qy 292 CTGTTTGAAACAGCGATCAGCCATTTGACCTGACCAAGTTCAAAAGCAAGAACGCCAG 351  
Db 1465 TATTATGAAAACAGCGACCGCCATTTGACTTAAACCAAGTTCACAAAGCAAGCCGCTGGT 1524  
Qy 352 TCCTCGCAAGCACATCTTGTATGTCCTCCACCTCAGAGCAGCTCTCTGACATGCC 411  
Db 1525 TCCAGCGTGGCTGTTCGGTGGCATCACCTCTGCGGGAGAGCGCACTCATGGACATCTCC 1584  
Qy 412 GACATGGTCAAAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTC 471  
Db 1585 GACATGGTGAACAACTCACAGCGCCCTGAGCCCAAGTCTCTCACGCCCTCCACAGT 1644  
Qy 472 CCCCCCATGAAGCTGGAATGATGTGAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531  
Db 1645 TCAGA---GAAATCCGATGCTGATGGCAGCAGCTTTGAGGAGCGCTTGAGCAGGTGCA 1701  
Qy 532 ACTTTGCAATAAGAAAGGCGGCGAGTCCCACTGGAATCTCTCAGCATCTTCTGATTTCA 591  
Db 1702 CCGGTCCCAAGAGAAAGGCGGCGAGTCCCACTGGAATCTCTCAGCATCTTCTCCTATCCTG 1761  
Qy 592 CAAGCCCATGTTGCTCTGAGCCTCTTCCAGCATCAGAGGGCAAAATACCTGCTGCTGAT 651  
Db 1762 CAGGCCAGTTCGCTCTGAGCTTGGGGAGACCAAGAGGCGAAGTATCATATGTCGAC 1821  
Qy 652 CTGGCCCAACAGAGCGTATGCAATCTCTAAGTTTACGGGATCTCTCAATGACCACTATC 711  
Db 1822 TTGGCCCGCAGGAGAGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACTATC 1881  
Qy 712 AGTCACTGGCTGGCCCAAGTCAAGTACCACTTAGGAAAACGGGCGGACAAAATTTCTG 771  
Db 1882 AGCCACTGGCTGGCCCAATGTGAAGTACCAGTTGAGGAGGACAGGGGGAAACGAAATTCCTA 1941  
Qy 772 AAAACATGGACAAAGGCGACCCATCTTTTATTCAGTGAAGTGTGCTCCAGTTTCA 831  
Db 1942 AAGAACTGGACAGAGGATCTCTGTTTCTTTTGAACGATTTGGCTCTCAGTTTCA 2001  
Qy 832 ACCCTTCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGACATG 891  
Db 2002 ACTGCTTCTACATATAAGTCAATTTGGAGACACACTTTGGGCTTCAGCTGAAGATCTC 2061  
Qy 892 ACCCGCTTGTCAAGTGGACCAAGAGTGGAGCAAGATCTCCGGGTATCTGCTG 951  
Db 2062 TCCAAGTGGCACT---CAATCAGATTCAAGAACAGCAGAAATGTTTCAAAATCTCTACC 2118  
Qy 952 GCTCAGAGTCTCCAGAAACAAATAGCTGCGAAGAGGACACAGACTCTTAAATTCAGTGT 1011  
Db 2119 AACAAATCTGGGCGCACTGGGGCCACCGAGGAGACTTGGGCTCCACATTCATATG 2178  
Qy 1012 AAGTTGTCTGTCGGACATTTGTGAGCAAAACATGCGGTAAACCTCACCTAAGCAAAACG 1071  
Db 2179 AAGCTCTCAACCGGACTTTTGGAGCAAGCAGCAGTCAAACTGACCTTAGTAAGACC 2238  
Qy 1072 CACAGCAAGTCAACCGAACCAATTCAGTTTGTAAACAGAGCTGGATGAAGATAGC 1129  
Db 2239 CACGGCAAGTCTCCCGAGGACCACTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 2296

RESULT 9

US-09-823-245A-409

; Sequence 409, Application US/09823245A  
; Publication No. US20020039760A1  
; GENERAL INFORMATION:  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakar  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6401  
; CURRENT APPLICATION NUMBER: US/09/823,245A  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/194,941  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 631  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 409  
; LENGTH: 3907  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 3870  
; OTHER INFORMATION: n = a,t,g, or c  
US-09-823-245A-409

Query Match 30.3%; Score 390.4; DB 9; Length 3907;  
Best Local Similarity 62.4%; Pred. No. 4.8e-111;  
Matches 673; Conservative 0; Mismatches 381; Indels 24; Gaps 3;  
Qy 52 AGCAATGGTGGCGCTCGCCCAACCCAGCCCGCCCTGCCATGTCATCAACCCACTCAGC 111  
Db 2425 AACAACTGGGATCATATGACCACTACCGGAGCCTTCTTTCATCAACCGCTGAGC 2484  
Qy 112 GCCCTGAGTCGCTCTGAAACAATCACTTGGGCAAGCCACGAGCCCTTGGCTCACT 171  
Db 2485 GCTTTGAGTCCATCATGAACACCCACCTGGGCAAGGTGTCCAAAGCCGCTG----- 2535  
Qy 172 TCCTGCTCCAGCCCAAGTTCAAGCACAAATTTCCATGTTTCCCAAGTCAATCTCAATGTC 231  
Db 2536 -----AGTCCCTCGCTGGACCCCGCTGGCGATGCTGTACAAAGTACGCAACAGCATG 2586  
Qy 232 ATGACAAAGCCGCTCTTGGTCTCCCAAGTCCAGCCAGCGGTGTCCAGGCGCTAC 291  
Db 2587 CTGGACAAGCCGCTGTATCCCGCCACCCCTGTGAAGAGCGCGATGCCATCGACCGCTAC 2646  
Qy 292 CTGTTTGAAACAGCGATCAGCCCAATTTGACCTGACCAAGTCCAAAGCAAGAACGCCAG 351  
Db 2647 TATTATGAAAACAGCGACCGCCATTTGACTTAAACCAAGTCCAAAGCAAGCCGCTGGT 2706  
Qy 352 TCCTGCGAAGCACAACTTTGTATGTCTCCCACTCAGAGCAGCGTCTGTCTGACATCGCC 411  
Db 2707 TCCAGCGTGGCTGATTCGGTGGCATCACCTCTGCGGAGAGCGCACTCATGACATCTCC 2766  
Qy 412 GACATGGTCAAAAGTCTCTCCCAAGCCACCCCAAGCCAGCCCTCTCTCCAGGGTC 471  
Db 2767 GACATGGTGAACAACTCACAGGCGGCTGAGCCCAAGTCTCTCACGCCCTCCACAGTT 2826  
Qy 472 CCCCCCATGAAGCTGGAATGATGTCTAGGCGCTTTGAGGATGTCTCCAGTGAAGTCTCA 531  
Db 2827 TCAGA---GAAAGTCCGATGCTGATGGCAGCAGCTTTGAGGAGCGCTTGGAGCGTGTCA 2883  
Qy 532 ACTTTGCAATAAGAAAGGCGGCGAGTCCAACTGGAATCTCTCAGCATCTTCTGATTTCA 591  
Db 2884 CCGGTCCCAAGAGAGAGGCGGCGAGTCCAACTGGAAACCCGCGAGCACCTTCTCTATCTCTG 2943  
Qy 592 CAAGCCCAAGTTCGCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGCTGCTGAT 651  
Db 2944 CAGGCCCAAGTTCGCTCGAGCTTGGGGAGAGCAACAGAGGGCAAGTATCATATGTCGGAC 3003

```
QY 652 CTGGGCCCAAGAGCGTATGCAAACTCTTAAGTTTAAGGGACTCTCAATGACCACTATC 711
Db 3004 TTGGGCCCCAGAGAGGGGTGCACATCTCGAAGTTTACTGGGCTCTCCATGACCACTC 3063
QY 712 AGTCACTGGCTGGCCAACTCAAGTACCAAGCTTTAGGAAAACGGGGGACAAAATTTCTG 771
Db 3064 AGCCACTGGCTGGCCAACTGTAAGTACCAAGTTGAGGAGCAGGGGGAACGAAATTCCTA 3123
QY 772 AAAACATGGACAAGGCCACCCCATCTTTTATTGGAGTGACCTGCGCTCCAGATTGAGA 831
Db 3124 AAGAACCTGGACACAGGGGATCCTGTTTCTTTTTCGAACGATTTGTGCTCTCAGTTTCA 3183
QY 832 ACCCTTCTACCTACATCAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGACATG 891
Db 3184 ACTGCTTCTACATACATAGTCACTTTGGAGACACACTTGGGCTTCAGCTGAAGATCTC 3243
QY 892 ACCCGCTTGTGCTGAGTGGACCAAGCAAGGTGGAGCAAGAGATCTCCCGGATATCGTCG 951
Db 3244 TCCAAAGCTGCCACT---CAATCAGATTCAAGAACAGCAGCAATGTTTCGAAAGTCTCTCACC 3300
QY 952 GCTCAGAGTCTCCAGAAACATAGCTGCGGAGAGGACACAGACTCTAAATTCAGTGT 1011
Db 3301 AACAAAACCTCTGGGCCCACTGGGGGCCACCGAGGAAGACTTGGGCTCCACATTTCCAATGT 3360
QY 1012 AAGTTGTGCTGTCGGACATTTGTGAGCAAAACATGCGGTAAACTCCACCTAAGCAAAACG 1071
Db 3361 AGCTCTGCAACCGGACTTTTGGAGCAAGCAGCAGTCAAACTGCACCTTTAGTAGACC 3420
QY 1072 CACAGCAAGTACCCGAAACCACTTACAGTTTGTAAACAGAGCTGGATGAAGAAATAGC 1129
Db 3421 CACGGCAAGTCTCCGAGGACCACCTGATCTATGTGACTGAGTTGGAGAAACAGTAGC 3478

RESULT 10
US-10-723-860-2953
; Sequence 2953, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2953
; LENGTH: 3692
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-2953

Query Match 21.1%; Score 272; DB 18; Length 3692;
Best Local Similarity 62.2%; Pred. No. 6.2e-74;
Matches 469; Conservative 0; Mismatches 270; Indels 15; Gaps 2;

QY 376 TCCCCACCTCAGAGCAGCTGTCTGACATCGCGGACATGGTCAAAGTCTCTCCCAAA 435
Db 2505 TCGCCGCTAGCGAGAAATGCTTGTGATATATCCGATATGCTGAAGACTTGACAGAG 2564
QY 436 GGCACACCCCAAGCCAGCTCTCTCCAGGGTCCCCCCCCATGAAGCTGGAAATGGAT 495
Db 2565 AGCCACACGTCAAAATCTCTCACTCTCTCCAGCATCTCC-----GAGAAGTCTGACATT 2618
QY 496 GTGAGGGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTGCATATAAAGAAAGCCGG 555
Db 2619 GACGGGGCCACTCTTGAGGAGGCTGAGGAGTGCACCCGCCAGGAAGGAGGCCGCGC 2678
```

```
QY 556 CAGTCAAACTGGAATCTCTCAGCATCTTCTGATTCTACAAGCCAGAGTTTGCCTCGAGCCTC 615
Db 2679 CAGTCAAACTGGAATCTCTCAGCATCTTCTGATTCTCCAAGCCAGAGTTTGCCTCGAGCCTC 2738
QY 616 TTCCAGACATCAGAGGGCAAAATACCTGCTGTGTGATCTGGGGCCCAAGAGAGGTATGCAA 675
Db 2739 CGGCAGACCTCAGAGGGGAAGTACATCATGTTCAGACCTTGAGCCCTGAGCGGGATGCAT 2798
QY 676 ATCTCTTAAGTTTACCGGACTCTCAATGACCATATCAGTCACTGGTGGCCCAACGTCAGAG 735
Db 2799 ATCTCTTAAGTTTACCGGACTCTCAATGACCATATCAGTCACTGGTGGCCCAACGTCAGAG 2858
QY 736 TACCAAGCTTAGGAAAACGGGGGAGCAAAATTTCTGAAAAACATGACAAAAGGCCACCC 795
Db 2859 TACCAAGCTTCGAAAGCAGGTGAGAACAAAGTTCTCTCAAAACCTTGGACACTGGCCACCC 2918
QY 796 ATCTTTTATTCAGTGACTGTCGCTCCAGGTTTCAGAAACCCCTTCTACCTACATCAGTCA 855
Db 2919 GTCTTCTTTTGAACGATTTGTCGTCCTCCAAATCAGGACTCTCTTCCACGTACATCAGTCA 2978
QY 856 TTAGAATCTCACTCGGTTTCCAAATGAAGGACATGACCCGCTTGTCACTGGACCAAGCAA 915
Db 2979 CTAGAGTCACATCTAGGCTTCCGGCTACGGGACTTATCCAAACTGTCCACCAACAGATT 3038
QY 916 AGCAAGGTGGAGCAAGAGATCTCCGGGTATCGTGGCTCAGAGGTCTCCAGAAACAAATA 975
Db 3039 AACA-----GTCAGATAGCACAAACCAAGTCACCGTCAGAAAAAATGGTACGCTCC 3089
QY 976 GTCGCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGTGTGTCGGACATTTGTG 1035
Db 3090 TCCCCCGAAGAGACCTGGGGACTTCTCTATCAGTGCAAAATTTCGCAATCGGACCTTTGCC 3149
QY 1036 ASCAAACATGCGGTAAACTCCACCTTAAGCAAAACGACAGCAAGTCACTCCCGAACACCAT 1095
Db 3150 AGCAAGCAGCTGTAAACCTTCACTTAGCAAAACACACGGGAATCTCCGGAAGACAC 3209
QY 1096 TCACAGTTTGTAAACAGACGTCGGATGAAGAAATAGC 1129
Db 3210 CTTCTGTATGTCTCTGAGTTAGAGAAGCAGTAGC 3243

RESULT 11
US-10-723-860-5900
; Sequence 5900, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; PRIOR FILING DATE: 2003-11-26
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5900
; LENGTH: 5065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (520)...(614)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4376)...(4392)
; OTHER INFORMATION: n is a, c, g, or t
US-10-723-860-5900
```

Query Match 21.1%; Score 272; DB 18; Length 5065;



; Sequence 7288, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods for Diagnosis of Soft Tissue Sarcoma, Compositions &  
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 7288  
; LENGTH: 5065  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (520)..(614)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4376)..(4392)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-723-860-7288

Query Match 21.1%; Score 272; DB 18; Length 5065;  
Best Local Similarity 62.2%; Pred. No. 7.1e-74;  
Matches 469; Conservative 0; Mismatches 270; Indels 15; Gaps 2;  
QY 376 TCCCGACCTCAGAGCAGCTCTGTCTGACATGCCGACATGTCAGAGTCTCCCAAA 435  
DB 2644 TCGCCCTACGCGAGAATGCTTGTGAGATATATCGGATATGCTGAAGAACTTGACAGAG 2703  
QY 436 GCCACCAACCCCAAGCAGCTCTCTCCAGGGTCCCGCCCATGAAGCTGGAATGGAT 495  
DB 2704 AGCCACAGTCAAAATCTCTCACTCTTCCAGATCTCC-----GAGAGTGTGACAT 2757  
QY 496 GTCCAGCGCTTTGAGATGCTCCAGTGAAGTCTCAACTTTGCATAAAAAGAGCCGG 555  
DB 2758 GACGGGGCACTCTGGAGGAGGTGAGGAGTGCAGCCGCCCGCAGAGGAGGCGCG 2817  
QY 556 CAGTCAACTGGAATCCTCAGCATCTTCTGATTTACAGCCCACTTTGCTCGAGCCTC 615  
DB 2818 CAGTCAAACTGGAACCCCGCAGCACTCTCTGATCTCCAGGCCAGTTTGGCGCCAGCCTC 2877  
QY 616 TTCCGACATCAGAGGCAAAATACCTGCTGTCTGATCTGGGCCCAAGAGCGTATGCAA 675  
DB 2878 CGGCAACCTCAGAGGGAAGTACATCATGTGAGACTGAGCCCCCAGAGCGGATGCAT 2937  
QY 676 ATCTTAAGTTTACGGACTCTCAATGACCACTATCAGTCACTGGCTGGCCCAACGTCAG 735  
DB 2938 ATCTCAGGTTTACCGGGCTGTCCATGACCACCATCAGCACTGGCTGGCCCAACGTCAG 2997  
QY 736 TACCAGCTTAGGAAGAGCGGGGACAAATTTCTGAAAAACATGACAAAGGCCACCCC 795  
DB 2998 TACCAGCTTCGAAGGACAGGTGGAACAAAGTTCTCTCAAAACATTTGGACACTGGCCACCCC 3057  
QY 796 ATCTTTTATTGAGTGTGCTCCAGTTCAGAACCCCTCTTACCTACATCAGTCTAC 855  
DB 3058 GTCTCTTTTGTAAACGATGTGGTGTCCCAATCAGAGCTCTTCCAGCTACATCAGTCTAC 3117  
QY 856 TTAGAAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGGTTGTGAGTGGACCAAGAA 915  
DB 3118 CTAGAGTCACTTAGGCTTCCGGCTAGCGGACTTATCCAAAGTGTCCACCGAAGATT 3177  
QY 916 AGCAAGGTGAGCAAGAGATCTCCCGGATTCGTCTCGCTCAGAGTCTCCAGAAACAATA 975  
DB 3178 AACA-----GTGAGTAGACAAACCAAGTCCCGTTCAGAAAAAATGGTGACGTCC 3228

QY 976 GTCCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGCTGTGCTCGGACATTTGTG 1035  
DB 3229 TCCCCCGAGGAAGACCTGGGACTTCTTATAGTCAAACTTTGCAATCGGACCTTTGCC 3288  
QY 1036 AGCAAAACATGCGGTAAACTCCACTTAAGCAAAAAGCAGCAAGTCAACCCGAAACCAT 1095  
DB 3289 AGCAAGCAGCGCTGTAAACTTCCACTTAGCAAAAACACACGGGAAATCTCCGGAAGACCAC 3348  
QY 1096 TCACAGTTTGTAAACAGACTGTGATGAAGATAGC 1129  
DB 3349 CTTCTGTATGCTCTGAGTTAGAGAGCAGTAGC 3382  
RESULT 14  
US-09-739-707-1  
; Sequence 1, Application US/09739707  
; Patent No. US20020142298A1  
; GENERAL INFORMATION:  
; APPLICANT: Chen, Hong  
; Friemer, Nelson, B.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR  
; THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC  
; DISORDERS  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/739,707  
; FILING DATE: 23-Mar-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/828,008  
; FILING DATE: 1997-07-30  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-071  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1144 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: Fsh 22  
; LOCATION: 1...282  
; OTHER INFORMATION:  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-739-707-1

Query Match 19.9%; Score 256.4; DB 9; Length 1144;  
Best Local Similarity 67.2%; Pred. No. 2.9e-69;  
Matches 437; Conservative 0; Mismatches 206; Indels 7; Gaps 5;  
QY 480 GAAGCTGGAATGAGTGTGAGCGCTTTTCAGAGTGTCTCAGTGAAGTCTCAACTTTGCA 539  
DB 63 GAAGTCCGATGCTGATGGCAGTAGCTTTGAGGAGGGTTGGAGCAGCTGTCCACCGTCCA 122  
QY 540 TAAAGAAAAGGCGCGCAGTCCCACTGGAATCCTCAGCATCTTCTGATTTCTACAAGCCCA 599







GenCore version 5.1.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:17:50 ; Search time 5451.8 Seconds  
(without alignments)  
8992.766 Million cell updates/sec

Title: US-08-731-499-5  
Perfect score: 1288  
Sequence: 1 GAGGCACGAGAGAGAA.....AAAAAAAAAAAAAAAAATTC 1288

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

1: gb\_est1: \*  
2: gb\_est2: \*  
3: gb\_hc: \*  
4: gb\_est3: \*  
5: gb\_est4: \*  
6: gb\_est5: \*  
7: gb\_est6: \*  
8: gb\_gssi: \*  
9: gb\_gss2: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1280.2	99.4	4149	3	HS0806854	BX640770 Homo sapi
2	793.6	61.6	1705	3	AK045693	AK045693 Mus muscu
3	682	53.0	727	1	AI889627	AI889627 wm59g03.x
4	677.4	52.6	680	6	CA449448	CA449448 UI-H-E11-
5	627	48.7	778	6	CB249889	CB249889 UI-M-EXO-
6	613.6	47.6	639	2	AW516191	AW516191 xt62d04.x
7	525.8	40.8	892	7	CN533555	CN533555 909277 MA
8	524.8	40.7	671	7	CN538481	CN538481 UI-M-HUO-
9	512.8	39.8	770	5	BU204906	BU204906 604157793
10	505	39.2	953	2	BF140337	BF140337 601787846
11	493	38.3	494	7	CR745711	CR745711 CR745711
12	475.4	36.9	477	1	AA873869	AA873869 oh79d06.s
13	465.2	36.1	686	7	CK773713	CK773713 962388 MA
14	438	34.0	1221	5	BQ715061	BQ715061 AGENCOURT
15	435.2	33.8	581	7	CK775937	CK775937 967078 MA
16	429	33.3	506	4	BM255904	BM255904 518125 MA
17	426	33.1	426	1	AA406466	AA406466 zv10c09.r
18	410.6	31.9	578	4	BG087988	BG087988 H3147C07-
19	406.8	31.6	492	1	AI593964	AI593964 vo99g12.y
20	393.6	30.6	2148	9	AY399842	AY399842 Mus muscu
21	393.6	30.6	3369	3	BC049795	BC049795 Mus muscu
22	393.6	30.6	3370	3	BC049799	BC049799 Mus muscu
23	390.4	30.3	2059	3	BC010679	BC010679 Homo sapi
24	390.4	30.3	2625	3	BC024152	BC024152 Homo sapi

25	389.4	30.2	2127	9	AY399840	AY399840 Homo sapi
26	382.6	29.7	417	7	HI6954	HI6954 ym34b09.81
27	381.4	29.6	1958	3	AK017514	AK017514 Mus muscu
28	379.2	29.4	458	7	HI2950	HI2950 y170h05.r1
29	378.6	29.4	529	2	AW821942	AW821942 uq35h05.y
30	377	29.3	450	7	CK388890	CK388890 L0935G05-
31	376.8	29.3	496	1	AA763786	AA763786 vo99g12.r
32	373	29.0	373	4	BM672622	BM672622 UI-B-CRO-
33	373	29.0	445	7	CK387442	CK387442 L0915G06-
34	369.2	28.7	430	4	BM105530	BM105530 509005 MA
35	369	28.6	514	7	HI6953	HI6953 ym34b09.r1
36	361	28.0	641	5	BP144926	BP144926 BP144926
37	359.6	27.9	505	2	BF193928	BF193928 245665 MA
38	356.8	27.7	486	4	EG070003	EG070003 H3083E07-
39	333.4	25.9	518	7	CN696106	CN696106 E0374D11-
40	331.8	25.8	578	7	CF739142	CF739142 UI-M-HDO-
41	330	25.6	547	5	BU102107	BU102107 PRODIG1D
42	329.8	25.6	841	7	CF242878	CF242878 AGENCOURT
43	327	25.4	488	4	BM288023	BM288023 528879 MA
44	319	24.8	2127	9	AY399841	AY399841 Pan trogl
45	318.8	24.8	953	6	BY1717304	BY1717304 BY1717304

#### ALIGNMENTS

RESULT 1  
HSM806854  
LOCUS HSM806854 4149 bp mRNA linear HTC 22-SEP-2004  
DEFINITION Homo sapiens mRNA; cDNA DKFZp686F07129 (from clone DKFZp686F07129).  
ACCESSION BX640770  
VERSION BX640770.1 GI:34364881  
KEYWORDS HTC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4149)  
AUTHORS Ansonge, W., Krieger, S., Regiert, T., Rittmueller, C., Schwager, B.,  
Mewes, H.W., Weil, B., Amid, C., Osanger, A., Fobo, G., Han, M. and  
Wiemann, S.  
CONSRSTM The German cDNA Consortium  
TITLE Direct Submission  
JOURNAL Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764  
Neuherberg, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;  
sequenced by EMBL (European Molecular Biology Laboratories,  
Heidelberg/Germany) within the cDNA sequencing consortium of the  
German Genome Project.  
This clone (DKFZp686F07129) is available at the RZPD Deutsches  
Resourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.  
Please contact RZPD for ordering:  
http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp686F07129  
Further information about the clone and the sequencing project is  
available at http://mips.gsf.de/projects/cdna/  
Location/Qualifiers  
1. 4149  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="RZPD:DKFZp686F07129Q"  
/db\_xref="taxon:9606"  
/clone="DKFZp686F07129"  
/tissue type="uterus"  
/clone lib="686 (synonym: hlccc3). Vector pSport1\_Sfi; host  
DH10B; sites SfiIA + SfiIB"  
/dev\_stages="adult"  
/note="chromosome 20 open reading frame 17 protein"  
1. 4149  
/gene="DKFZp686F07129"  
888. 3992  
/gene="DKFZp686F07129"  
/codon\_start=1

```

/product="hypothetical protein"
/protein_id="CAE45871.1"
/db_xref="GI:34364882"
/db_xref="UniProt/TREMBL_Q6N003"
/translation="MPRRKQAPKRAAGYAQAEQLKEBEEIKBEEBEEEDSGVAQLQG
GNDTGDELETPEQKGCFSYONSPGSHLSNOAENESLLSDASDQVSDIKSVCGRD
ASDKAHTHVLSPNEAHNCMDKWTAVYANILSDSYWSGLGFLKLSNSERRNCDTRNG
SNKSDFDHQLDLSKLOQLPSRSVSKPSLFSYQCRQSSKMGCTVTFGASRFRCR
QCSATVDLVELTAPKMHNETGHYDDNRKDKLRPTYSKPRKRAFOFMDKEDAOKVLK
CMFGSDSPSLDLSVHMLIKHYOKVPLKEPVPTLSKMWTPAKRVPDVRNRPSPD
STTGSFADSPSSQKQANLQSSNNRYGNGASYTWQPEACKSQILKMECCSHDTL
QDLTHMVTYVLTGTSKSKGKQVLDPLAVERKQSLSEAPNSDLSLAPKPSNSA
SDCTATTHKESKEKPEETSDEKVKVKSDEYEDPLQKPLDPTIKYLYLEEDLED
GSKGGDILKLSLSTVTTAINKAQNGAPSWAYPSIHAAYQLSEBGTPLPLPMGSOVLQ
IRPNLTKLRPTAPKMWPLVSMPTHLAPYQVKKSEDEKDAVKECKGKESPHREAS
SPSHSGDSFRSETPPEAKYTELGPLEKEEKLMEGSEKPEQPLEPTSAISNGCAL
ANHPALPCINLPSALQSVLNHNLKATPLKSPSCSPSSSTISMFHKSNNLVMDKP
VLSPPASTRASVSRYPFLFENSDDPDLTKSKKAESSQAQSCMPPOKHALSDADM
VKVLPKATTPKPASSSRVPMKLEMDRFPEDVSEVSTLHKRKGQSNWNPQHLLIL
QAQFASLFTSEGKYLSDLGPOERMQISKEFTGLSMVTISHLANVKYQLRKTGTTK
FPKMDKGHPITFYCSDCAQSFRTPTIYISHLESHLGFOKMDTRLSVDQGSKVEQIS
RVSSAQRSPETTAABEDTSPKPKLCCRTFVSKHAVKLHLKTHSKSPHHSQFVTD
VDEE"

```

ORIGIN

Query Match	99.4%; Score 1280.2; DB 3; Length 4149;				
Best Local Similarity	99.8%; Pred. No. 0;				
Matches 1282; Conservative	0; Mismatches 3; Indels 0; Gaps 0;				
QY 1	GAGGGCAGCGAGAGAGAAACCCAGCCCTGAGGCCACATCTCTCTGAGCAATGGG 60				
DB 2865	GAGGGCAGCGAGAGAGAAACCCAGCCCTGAGGCCACATCTCTCTGAGCAATGGG 2924				
QY 61	TGCGCCTCGCCACACACGCGCCGCTGCCATGATCAACCCACTCAGCGCCCTGCAG 120				
DB 2925	TGTGCGCCTCGCCACACACGCGCCGCTGCCATGATCAACCCACTCAGCGCCCTGCAG 2984				
QY 121	TCCGTCTGAAACATCACTTGGGCAAAAGCCAGGAGCCCTTGGCGCTCACCTCTCTGCTCG 180				
DB 2985	TCCGTCTGAAACATCACTTGGGCAAAAGCCAGGAGCCCTTGGCGCTCACCTCTCTGCTCG 3044				
QY 181	AGCCCAAGTTCAAGCACAATTTCCATGTGTTCCACAAGTCGAATCTCAATGTGATGGACAAG 240				
DB 3045	AGCCCAAGTTCAAGCACAATTTCCATGTGTTCCACAAGTCGAATCTCAATGTGATGGACAAG 3104				
QY 241	CGGTCCTGAGTCTGCTCCCAAGAGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGAG 300				
DB 3105	CGGTCCTGAGTCTGCTCCCAAGAGTCAGCCAGCGTGTCCAGGCGCTACCTGTTTGAG 3164				
QY 301	AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAACCCGAGTCTCTCGCAA 360				
DB 3165	AACAGCGATCAGCCCATTTGACCTGACCAAGTCCAAAAGCAAGAACCCGAGTCTCTCGCAA 3224				
QY 361	GCACAATTTGTATGTGTCCTCCACCTCAGAACGCGTCTGTGTGACATCGCCGACATGGTC 420				
DB 3225	GCACAATTTGTATGTGTCCTCCACCTCAGAACGCGTCTGTGTGACATCGCCGACATGGTC 3284				
QY 421	AAAGTCTCCCAAGAGCCACCCCAAGCCAGCCTCTCTCCAGGCTCCCGCCCATG 480				
DB 3285	AAAGTCTCCCAAGAGCCACCCCAAGCCAGCCTCTCTCCAGGCTCCCGCCCATG 3344				
QY 481	AAAGTCTCCCAAGAGCCACCCCAAGCCAGCCTCTCTCCAGGCTCCCGCCCATG 3404				
DB 3345	AAAGTCTCCCAAGAGCCACCCCAAGCCAGCCTCTCTCCAGGCTCCCGCCCATG 3464				
QY 541	AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCAGCATCTTCTGATTTACAAAGCCAG 600				
DB 3405	AAAAGAAAAGCGCGGAGTCCAACTGGAATCTCAGCATCTTCTGATTTACAAAGCCAG 3464				
QY 601	TTTGCCCTCGAGCCTCTCCAGACATCAGAGGCAAAATACCTGCTGTGATCTGGGCCCA 660				
DB 3465	TTTGCCCTCGAGCCTCTCCAGACATCAGAGGCAAAATACCTGCTGTGATCTGGGCCCA 3524				

QY 661	CAAGACGCTGATGCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 720				
DB 3525	CAAGACGCTGATGCAAAATCTTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGG 3584				
QY 721	CTGGCCAAAGCTCAAGTACCAAGCTTAGGAAAACGGGGGGGACAAAATTTCTGAAAAACATG 780				
DB 3585	CTGGCCAAAGCTCAAGTACCAAGCTTAGGAAAACGGGGGGGACAAAATTTCTGAAAAACATG 3644				
QY 781	GACAAAGGCCACCCCATCTTTTATTTCAGTGTACTGTGCTCCAGTTTCAGAAACCCCTTCT 840				
DB 3645	GACAAAGGCCACCCCATCTTTTATTTCAGTGTACTGTGCTCCAGTTTCAGAAACCCCTTCT 3704				
QY 841	ACCTACATCAGTCACTTAGAATCTCACTCGGGTTTCCAAATGAAGACATGACCCGCTTG 900				
DB 3705	ACCTACATCAGTCACTTAGAATCTCACTCGGGTTTCCAAATGAAGACATGACCCGCTTG 3764				
QY 901	TCAGTGGACCAAGAACAGTGGAGCAGAGATCTCCCGGTATCTCGGCTCAGAGG 960				
DB 3765	TCAGTGGACCAAGAACAGTGGAGCAGAGATCTCCCGGTATCTCGGCTCAGAGG 3824				
QY 961	TCCTCAGAAAACATAGTCTGCCGAAGAGGACACAGACTCTTAAATTCAGTGTAAAGTTGTGC 1020				
DB 3825	TCCTCAGAAAACATAGTCTGCCGAAGAGGACACAGACTCTTAAATTCAGTGTAAAGTTGTGC 3884				
QY 1021	TGTCGGACATTTGTGAGCAACATCGGTAAACCTCCACTAAGCAAAAACGACAGCAAG 1080				
DB 3885	TGTCGGACATTTGTGAGCAACATCGGTAAACCTCCACTAAGCAAAAACGACAGCAAG 3944				
QY 1081	TCACCGGAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGAATAGCTCTGCGAGGCG 1140				
DB 3945	TCACCGGAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGAATAGCTCTGCGAGGCG 4004				
QY 1141	AATGCTTAGTTTCCACTTTCCAGCTGATCCCTCACACTGAACCTTCTTCTGTTGCA 1200				
DB 4005	AATGCTTAGTTTCCACTTTCCAGCTGATCCCTCACACTGAACCTTCTTCTGTTGCA 4064				
QY 1201	CCATCTGCTCTGACATTTGAACTCATTGAACTCTCTCTGACACCTGCTGCTGAGAGA 1260				
DB 4065	CCATCTGCTCTGACATTTGAACTCATTGAACTCTCTCTGACACCTGCTGCTGAGAGA 4124				
QY 1261	CTGCCAAAAAATAAAAAAAAAAAAAA 1285				
DB 4125	CTGCCAAAAAATAAAAAAAAAAAAAA 4149				
RESULT 2					
AK045693	1705 bp mRNA linear HTC 03-APR-2004				
LOCUS	Mus musculus adult male corpora quadrigemina cDNA, RIKEN				
DEFINITION	full-length enriched library, clone:B230304M06 product:TEASHIRT 2 (FRAGMENT) homolog [Mus musculus], full insert sequence.				
ACCESSION	AK045693				
VERSION	AK045693.1 GI:26337546				
KEYWORDS	HTC; CAP trapper.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20499374				
PUBMED	11042159				
REFERENCE	3				

## AUTHORS

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaki, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Iihikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861

## REFERENCE

## AUTHORS

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

## REFERENCE

## AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs  
Nature 420, 563-573 (2002)  
6 (bases 1 to 1705)

## REFERENCE

## AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hasegawa, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submission

## REFERENCE

## JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.jp/

URL: http://fantom.gsc.riken.jp/

## FEATURES

## source

Location/Qualifiers  
1..1705  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:B230304M06"  
/db\_xref="taxon:10090"  
/clone="B230304M06"  
/sex="male"  
/tissue\_type="corpora quadrigenina"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"  
416..1570  
/note="unnamed protein product; TEASHIRT 2 (FRAGMENT) homolog [Mus musculus] (SPTR|Q9JUL71, evidence: FASTA, 98.6%id, 6.2%length, match=207)  
putative)  
/codon\_start=1  
/protein\_id="BAC32459.1"  
/db\_xref="GI:26337547"  
/translation="MKSLKQCGKSPHEPHEATSFQSGESFSKIEPPSPSRKABPCPL"

## CDS

ORIGIN  
Query Match 61.6%; Score 793.6; DB 3; Length 1705;  
Best Local Similarity 87.4%; Pred. No. 2.4e-206;  
Matches 881; Conservative 0; Mismatches 124; Indels 3; Gaps 1;  
QY 11 AGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGGCGCCCTCG 70  
DB 567 AGAAGGAGAAACCCAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGGCGCCCTCG 626  
QY 71 CCAACACGCGCCCGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGGCGCCCTCG 130  
DB 627 CTAACACGCGCCCGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGGCGCCCTCG 686  
QY 131 ACAATCACTTGGGAAAGCCAGGAGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGG 190  
DB 687 ACAACACGCGCCCGCCCTGGAGCCACATCTGCTCTGAGCAATGGGTGGCGCCCTCG 746  
QY 191 CAAGCAATTTCCATGTTTCCACAGTCGATCTCAATGTCTATGAGCAAGCGGTCTTGA 250  
DB 747 CAAGCAATTTCCATGTTTCCACAGTCGATCTCAATGTCTATGAGCAAGCGGTCTTGA 806  
QY 251 GTCCTGCTCCACAAAGGTCA---GCCAGCGTGTCCAGCGGTCTGAGCAAGCGGTCTTGA 307  
DB 807 GTCCACCTCCACAGCGCGCGCGCGCGCGGTGGGAGACATCTGCTTGGAGACACCG 866  
QY 308 ATCAGCCCAATGACTGACCAAGTCCAAAGCAAGAAAGCGAGTCTTGGCAAGCAAT 367  
DB 867 ACCAGCCCAATGACTGACCAAGTCCAAAGCAAGAAAGCGAGTCTTGGCAAGCAAT 926  
QY 368 CTTGTATGTCCCGACCTCAGAGCAAGTCTGCTGTGATCGCCGACATGTCAGAGTCC 427  
DB 927 CTTGTATGTCCCGACCTCAGAGCAAGTCTGCTGTGATCGCCGACATGTCAGAGTCC 986  
QY 428 TCCCAAGGACCAACCCCAAGCGAGTCTGCTGTGATCGCCGACATGTCAGAGTCC 487  
DB 987 TCCCAAGGACCAACCCCAAGCGAGTCTGCTGTGATCGCCGACATGTCAGAGTCC 1046  
QY 488 AAATGGATGTCCAGCGCTTTCAGGATGTCCTCAGTGAAGTCTCAATTTTGCATAAAGAA 547  
DB 1047 AAATGGATGTCCAGCGCTTTCAGGATGTCCTCAGTGAAGTCTCAATTTTGCATAAAGAA 1106  
QY 548 AAGCGCGGAGTCCCACTGGAATCTCAGATCTTCTGATTTCTACAGCCCGAGTTGGCT 507  
DB 1107 AAGCGCGGAGTCCCACTGGAATCTCAGATCTTCTGATTTCTACAGCCCGAGTTGGCT 1166  
QY 608 CGAGCTCTTCAGACATCAGAGGCAATACCTGCTGTGATCTGAGCCGCCCAAGAGC 667  
DB 1167 CGAGCTCTTCAGACATCAGAGGCAATACCTGCTGTGATCTGAGCCGCCCAAGAGC 1226  
QY 668 GTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGTGGCA 727  
DB 1227 GGATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGTGGCA 1286  
QY 728 AGTCAAGTACAGCTTAGGAAAACGGCGGAGCAAAATTTCTGAAAACATGGCAAG 787  
DB 1287 AGTCAAGTACAGCTTAGGAAAACGGCGGAGCAAAATTTCTGAAAACATGGCAAG 1346  
QY 788 GCCACCCATCTTTATTTGAGTGTGCTGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 847  
DB 1347 GGCACCCATCTTTATTTGAGTGTGCTGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 1406  
QY 848 TCAGTCACTTAGAATCTCAGCTGGGTTTCCAAATGAAGGACATGACCCCGTCTTGTGAGTGG 907  
DB 1407 TCAGTCACTTAGAATCTCAGCTGGGTTTCCAAATGAAGGACATGACCCCGTCTTGTGAGTGG 1466  
QY 908 ACCAGCAAGGAGGTGAGCAAGATCTCTCCCGGTTATCGTCCGCTCAGAGGTCTCCAG 967

```
Db 1467 ACCAGCAAGAGGTCGAGCAGAGATCTCCGAGTGTGTCGCTCAGAGGCTCTCCAG 1526
QY 968 AAACAATAGTGTCCGGAAGAGGACACAGACTCTAAATTCAGTCTAAGT 1015
Db 1527 AAACAATAGTGTCCGGAAGAGGACACAGACTCTAAATTCAGTCTAAGT 1574

RESULT 3
AI889627/c
LOCUS 727 bp mRNA linear EST 07-MAR-2000
DEFINITION wm5903.x1 NCI CGAP Ut2 Homo sapiens cDNA clone IMAGE:2440276 3'
similar to TR:060534 O60534 ANTIGEN NY-CO-33.; mRNA sequence.
ACCESSION AI889627
VERSION AI889627.1 GI:5594791
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 727)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 1472 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 412.
Location/Qualifiers
1..727
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2440276"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI-CGAP_Ut2"
/notes="Organ: uterus; Vector: pCMV-SPORT6; Site 1: SalI;
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

ORIGIN
Query Match 53.0%; Score 682; DB 1; Length 727;
Best Local Similarity 97.7%; Pred. No. 8.7e-176;
Matches 710; Conservative 0; Mismatches 15; Indels 2; Gaps 2;

QY 544 AGAAAGGCGGCGAGTCGCACTCGAATCTTCAGCA-TCTTCTGATTTCTAC-AAGCCAGT 601
Db 727 AGAAAGGCGGCGAGTCCACTCGAATCGTCAGCATTTCTGATTGNAACAAGCCAGT 668
QY 602 TTGCTTCGAGCTCTTCCAGACATCAGAGGCAATACCTGCTGTCGATCTGGGCCAC 661
Db 667 TGGCTTCGAGCTCTTCCAGACATCAGAGGCAATACCTGCTGTCGATCTGGGCCAC 608
QY 662 AAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGC 721
Db 607 AAGAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACTATCAGTCACTGGC 548
QY 722 TGGCCAAAGTCAGTACAGTCTAGGAAACCGGCGGGAACAAATTTCTGAAACAATGG 781
Db 547 TGGCCAAAGTCAGTACAGTCTAGGAAACCGGCGGGAACAAATTTCTGAAACAATGG 488
```

```
QY 782 ACAAGGCGCACCCCATCTTTTATTGTCAGTGAAGTGTGCTCCAGTTTCAGAACCCCTTCTA 841
Db 487 ACAAGGCGCACCCCATCTTTTATTGTCAGTGAAGTGTGCTCCAGTTTCAGAACCCCTTCTA 428
QY 842 CTTACATAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGTTGT 901
Db 427 CTTACATAGTCACTTAGAATCTCACTGGGTTTCCAAATGAAGGACATGACCCGTTGT 368
QY 902 CAGTGACAGCAAAAGCAGGTGGAGCAGAGATCTCCGGGTATCGTCGGCTCAGAGT 961
Db 367 CAGTGACAGCAAAAGCAGGTGGAGCAGAGATCTCCGGGTATCGTCGGCTCAGAGT 308
QY 962 CTCGAGAAACAATAGTGTGCGGAAGGACACAGACTCTAAATTCAGTGAAGTTGTGCT 1021
Db 307 CTCGAGAAACAATAGTGTGCGGAAGGACACAGACTCTAAATTCAGTGAAGTTGTGCT 248
QY 1022 GTGCGACATTTGTGAGCAAAACATGCGGTAAATCTCCACTTAAGCAAAACGACAGCAAGT 1081
Db 247 GTGCGACATTTGTGAGCAAAACATGCGGTAAATCTCCACTTAAGCAAAACGACAGCAAGT 188
QY 1082 CACCCGACACCATTCACAGATTTGTAACAGAGCTGGATGAAGATAGCTCTGCAGACCA 1141
Db 187 CACCCGACACCATTCACAGATTTGTAACAGAGCTGGATGAAGATAGCTCTGCAGACCA 128
QY 1142 ATGCTCTTAGTTCCTCACTTTCCAGCTTGGATCCCTCACAAGCTTCTTCTGTTGCAC 1201
Db 127 ATGCTCTTAGTTCCTCACTTTCCAGCTTGGATCCCTCACAAGCTTCTTCTGTTGCAC 68
QY 1202 CATCTGCTTCTGACATTTGAATTCATTTGAATTCCTCTGACACCTGCTGAGAAAGAC 1261
Db 67 CATCTGCTTCTGACATTTGAATTCATTTGAATTCCTCTGACACCTGCTGAGAAAGAC 8
QY 1262 TGCCTAAA 1268
Db 7 TGCCTAAA 1

RESULT 4
LOCUS CA449448 680 bp mRNA linear EST 08-NOV-2002
DEFINITION UI-H-E11-ayt-h-09-0-UI.s1 NCI CGAP E11 Homo sapiens cDNA clone
UI-H-E11-ayt-h-09-0-UI 3', mRNA sequence.
ACCESSION CA449448
VERSION CA449448.1 GI:24813868
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 680)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1..680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-E11-ayt-h-09-0-UI"
/tissue_type="Chondrosarcoma"
/dev_stage="Adult"
```

```
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_E11"
/notes="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
with a modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI CGAP E11 is a normalized cDNA library containing the
following tissue(s): Chondrosarcoma. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT7T3-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
ACACTTGAC.
TAG_TISSUE=chondrosarcoma
TAG_LIB=UI-H-E11
TAG_SEQ=ACACTTGAC"

ORIGIN
Query Match 52.6%; Score 677.4; DB 6; Length 680;
Best Local Similarity 99.7%; Pred. No. 1.6e-174;
Matches 678; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 604 GCCTGAGCCTTTCAGACATCAGAGGCAATACCTGCTCTGATCTGGGCCCAAA 663
DB 680 GCCTGAGCCTTTCAGACATCAGAGGCAATACCTGCTCTGATCTGGGCCCAAA 621

QY 664 GAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACCTATCAGTGGGCTG 723
DB 620 GAGCGTATGCAATCTCTAAGTTTACGGGACTCTCAATGACCACCTATCAGTGGGCTG 561

QY 724 GCAACGTCACAGTACCAGCTTAGGAAAACGGCGGACAAAATTTCTGAAAACATGGAC 783
DB 560 GCACAGTCAGTACAGCTTAGGAAAACGGCGGACAAAATTTCTGAAAACATGGAC 501

QY 784 AAAGGCCACCCCATCTTTTATTCAGTGTAGTGTGCTCCAGTTCAGAACCCCTTCTACC 843
DB 500 AAAGGCCACCCCATCTTTTATTCAGTGTAGTGTGCTCCAGTTCAGAACCCCTTCTACC 441

QY 844 TACATCAGTCACTAGAACTCACCCTGGTTTCCAAATGAAGACATGACCCGCTTGCA 903
DB 440 TACATCAGTCACTAGAACTCACCCTGGTTTCCAAATGAAGACATGACCCGCTTGCA 381

QY 904 GTGGACCAAGCAAGAGTGGAGCAAGAGATCTCCCGGATTCGTCGCTCAGAGGCT 963
DB 380 GTGGACCAAGCAAGAGTGGAGCAAGAGATCTCCCGGATTCGTCGCTCAGAGGCT 321

QY 964 CCAGAAACAATAGTGTCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGCTGT 1023
DB 320 CCAGAAACAATAGTGTCCGAAGAGGACACAGACTCTAAATTCAGTGTAAAGTTGCTGT 261

QY 1024 CGGACATTTGTGAGCAAAATCGGTAAATCTCCACCTAAGCAAAACGACGACAGTCA 1083
DB 260 CGGACATTTGTGAGCAAAATCGGTAAATCTCCACCTAAGCAAAACGACGACAGTCA 201

QY 1084 CCGAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGATAGCTCTGCAGGACGAAT 1143
DB 200 CCGAACACCATTCACAGTTTGTAAACAGACGTGGATGAAGATAGCTCTGCAGGACGAAT 141

QY 1144 GCCTTAGTTTCCACTTTCAGAGCTGGATCCCTTCCACTGAAACCCCTTCTCGTTCACCA 1203
DB 140 GCCTTAGTTTCCACTTTCAGAGCTGGATCCCTTCCACTGAAACCCCTTCTCGTTCACCA 81

QY 1204 TCTGTTCTGACATTTGAATCTATTTGAATCTCTCTGACACCCCTGGCTCTGAGAGACTG 1263
DB 80 TCTGTTCTGACATTTGAATCTATTTGAATCTCTCTGACACCCCTGGCTCTGAGAGACTG 21

QY 1264 CCAAAAAAAAAAAAAAAAAA 1283
DB 20 CCAAAAAAAAAAAAAAAAAA 1
```

RESULT 5  
CB249889  
LOCUS  
DEFINITION  
CB249889  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

CB249889 778 bp mRNA linear EST 15-JUL-2003  
UI-M-EXO-byo-k-23-0-UI-r1 NIH BMAP\_EXO Mus musculus cDNA clone  
IMAGE:5720974 5', mRNA sequence.  
CB249889.1 GI:28389782  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 778)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pYX-5.  
Location/Qualifiers  
1. 778  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5720974"  
/tissue\_type="whole brain"  
/dev\_stages="embryo 15.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP\_EXO"  
/note="Organ: brain; Vector: pYX-Rac; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction.  
ligated with EcoR I adaptor, digested with Not I, and then  
cloned directionally into pYX-Rac vector. The library tag  
sequence located between the Not I site and the polyA  
tail, is GTCCGTGGAA. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP). 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institutes of Mental Health  
(NIMH), Hemin Chin, Ph.D., program coordinator."

Query Match 48.7%; Score 627; DB 6; Length 778;  
Best Local Similarity 88.9%; Pred. No. 1.1e-160;  
Matches 675; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 257 COTCACAAGGTTCAGCCAGCGTGTCCAGCGCTACTGTTTGAGAACAGCGATCAGCCCA 316  
DB 8 CCACACGCGCGCGCCAGCGTGGCCAGACACTACTGTTTGAGAACACCGACAGCCCA 67

QY 317 TTGACCTACCAAGTCCAAAGCAAGACGAGTCTCTCGAAGACACATCTTGATGT 376  
DB 68 TTGACCTACCAAGTCCAAAGCAAGACGAGTCTCTCGAAGACACATCTTGATGT 127

QY 377 CCCCACCTCAGAACGACGCTCTGTGACATCGCGACATGCTCAAGTCTCCCAAG 436  
DB 128 CCCCACCTCAGAACGACGCTCTGTGTGACATGCTGTGATATGTTCAAGTCTCTCCCAAG 187



REFERENCE 1 (bases 1 to 892)  
AUTHORS Smith,T.P.L., Frelking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,  
Nonneman,D.J., Wray,J.E. and Keele,J.W.  
TITLE Porcine EST collection using a normalized library constructed from  
embryos representing early developmental stages  
JOURNAL Unpublished (2003)  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smithemail.marc.usda.gov  
Single pass sequencing. Bases called with phred v0.020425.c and  
trimmed with the aid of the trim\_alt option. Vector identified with  
cross\_match v0.990329.  
Plate: TMW8025 row: A column: 10  
Seq primer: GTAATACGACTCAGCTATAGG.  
FEATURES  
source  
Location/Qualifiers  
1..892  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/clone\_lib="MARC 4PIG"  
/note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;  
Library made with combined RNA from day-10, day-13,  
day-15, day-25, and day-30 whole embryos."  
ORIGIN  
Query Match 40.8%; Score 525.8; DB 7; Length 892;  
Best Local Similarity 90.8%; Pred. No. 7.3e-133;  
Matches 560; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
Qy 521 GTGAAGTCTCACTTTGCATATAAAGAAAGGCGGCGAGTCCAACTGGAAATCTCAGCATC 580  
Db 1 GCGAGGCTCGACCTTCATATAAAGAAAGGCGGCGAGTCCAACTGGAAATCTCAGCATC 60  
Qy 581 TTCTGATTTCTAAGCCAGCTTTGCTCGAGCTCTTCCAGACATCAGAGGCGCAATACC 640  
Db 61 TTCTGATCTCGAGGCTCAGTTTGTCTTCAGGCTCTTCCAGACCTCGGAGGCGCAATACC 120  
Qy 641 TGCTGTCTGATCTGGGCGCCAGAGGCTATGCAATCTCTAAGTTTACGGGACTCTCAA 700  
Db 121 TGCTGTCTGACCTGGGTCGCCAGGAGCGAATGCAGATCTCGAAGTTTCAGGACTCTCAA 180  
Qy 701 TGACCACTATCAGTCACTGGCTGGCCAACTCAAGTACCAGCTTAGGAAAGCGGCGGGA 760  
Db 181 TGACCACTATCAGCACCCTGGCTGGCCAACTCAAGTACCAGCTTAGGAAAGCGGCGGGA 240  
Qy 761 CAAATTTCTGAAAAACATGCAAAAGGCGCCATCTTTTATTCAGTCACTGTGCT 820  
Db 241 CGAATTTCTGAAAAACATGCAAAAGGCGCCATCTTTTATTCAGTCACTGTGCT 300  
Qy 821 CCCAGTTGAGAACCCCTTCTACCTACATCAGTCACTTAGAATCTCACTGGGTTCCAAA 880  
Db 301 CCCAGTTCCGAAAGCGCTTCGACCTACATCAGCCTTAGAATCTCACTAGGCTTCCAAA 360  
Qy 881 TGAAGGATGATGACCGCTTGTCTGAGTGGCCAGCAAGCAAGTGGGCAAGAGATCTGCC 940  
Db 361 TGAAGGATGATGACCGCTTGTCTGAGTGGCCAGCAAGCAAGTGGGCAAGAGATCTGCC 420  
Qy 941 GGGTATCGTTCGCTCAGAGGCTTCCAGAAACAAATAGCTGCCAGAGGACACAGACTCTA 1000  
Db 421 GGGTATCGTTCGCTCAGAGGCTTCCAGAAACAAATAGCTGCCAGAGGACACAGACTCTA 480  
Qy 1001 AATCAAGTGTAAAGTTGTCTGCGGCAATTTGTGAGCAAAACATCGCGGTAAACTCCACC 1060  
Db 481 AATCAAGTGTAAAGTTGTCTGCGGCAATTTGTGAGCAAAACATCGCGGTAAACTCCACC 540  
Qy 1061 TAAGCAAAACGACAGCAAGTCAACCGGACCACTTACAGTTCGTACAGACGCTGGATG 1120  
Db 541 TAAGCAAAACGACAGCAAGTCAACCGGACCACTTACAGTTCGTACAGACGCTGGATG 600

Qy 1121 AAGAAATAGCTCTGCAGG 1137  
Db 601 AGGAATRACTCTGCAGG 617  
RESULT 8  
CNS38481 671 bp mRNA linear EST 29-APR-2004  
LOCUS UI-M-HUO-cqs-i-15-0-UI.r1 NIH BMAP\_HUO Mus musculus cDNA clone  
DEFINITION IMAGE:30667886 5', mRNA sequence.  
ACCESSION CNS38481  
VERSION CNS38481.1 GI:46866637  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 671)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. James Lin University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Distribution information can be found at  
http://genome.uiowa.edu/distribution/mouseefl.html  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)  
Seq primer: pYX-5.  
FEATURES  
source  
Location/Qualifiers  
1..671  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30667886"  
/tissue\_type="whole eye"  
/dev\_stages="newborn ( 1, 5, 15 days )"  
/lab\_host="DH10B (T1 phage resistant)"  
/clone\_lib="NIH BMAP\_HUO"  
/note="Organ: Eye; Vector: pYX-Asc; Site 1: EcoR I;  
Site 2: Not I; The library was constructed according  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806.  
1996. Denatured RNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with oligo-dT  
primer containing a Not I site. Double strand cDNA was  
size selected according to mRNA size fraction, ligated  
with EcoR I adaptor, digested with NotI and then cloned  
directionally into pYX-Asc vector. The library tag  
sequence located between the Not I site and the polyA tail  
is AATAATATACG. This library was created for the University  
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the  
Developing Mouse Nervous System', supported by National  
Institute of Mental Health (NIMH)."  
ORIGIN  
Query Match 40.7%; Score 524.8; DB 7; Length 671;  
Best Local Similarity 91.4%; Pred. No. 1.3e-132;  
Matches 556; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
Qy 530 CAACCTTTCATAAAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTC 589  
Db 1 CCACCTTTCATAAAGAAAGCGCGGAGTCCAACTGGAATCCTCAGCATCTTCTGATTC 60  
Qy 590 TACAAGCCCGAGTTTGCCTTCGAGCCTCTTCCAGACATCAGAGGCGCAATACCTGCTGTG 649  
Db 61 TCGAAGCTCAGTTTGCCTTCGAGCCTCTTCCAGACCTCAGAGGCGCAATACCTGCTGTG 120  
Qy 650 ATCTGGGCGCCAAAGAGCGGTATGCAATCTCTTAAGTTTACGGGACTCTCAATGACCACTA 709



121	Db	 ACCTGGGCCCCCAAGAGCGGATGCCAAATCTCCAAGTTTCA	CGGACTCTCCATGACCCAA	180
710	Qy	 TCAGTCACTGGCTGGCCCAAGTCAAGTACAGCTTTAGGAAAA	CGGGCGGACAAAAATTC	769
181	Db	 TCAGCCACTGGCTGGCTTAAGTCMAATACAGCTTAGGAAAA	CAGGTGGGACAAAGTTCC	240
770	Qy	 TGAAAAACATGGACAAAGGCCACCCCATCTTTTATATGCA	TGTAAGTCTCCAGTTCA	829
241	Db	 TGAAAAACATGGACAAAGGGCACCCCATCTTTTACTGCA	TGTAAGTCTCCAGTTCA	300
830	Qy	 GAAACCCCTTCTACTACATCAGTCACTTAGAACTCACT	CGGTTTCCAAATGAAAGACA	889
301	Db	 GAAACCCCTTCTACTACATCAGGCACCTAGAGTCTCAC	TGGGCTTCCAAATGAAAGACA	360
890	Qy	 TGACCCCGTTGTTCAGTGGACACGAAAGCAAGGTGGAG	CAAGATCTCCCGGTATCGT	949
361	Db	 TGACCCGGATGGCGGCTGACACGAAAGCAAGGTGGAG	CAGGAGATCTCCGAGTGTCT	420
950	Qy	 CGGCTCAGAGGTCTCCAGAAACAATAGCTGCGGAAGAG	GACACAGACTCTAAATTCAGT	1009
421	Db	 CGGCTCAGAGGTCTCCAGAAACAATAGCTGCGGAAGAG	GACACAGACTCTAAATTCAGT	480
1010	Qy	 GTAAGTTGCTGTCCGACATTTGTGAGCAAAACATGCG	GTAAACTCCACTAAGCAAAA	1069
481	Db	 GTAAGTTATGCCGTGCGACATTTGTGAGCAAAACAC	GCAAGTAAACTCCACTAAGCAAAA	540
1070	Qy	 CGCACAGCAAGTCAACCCGAACACCATTCACAGTTTGT	TAAACAGAGCTGGATGAAGAAATAGC	1129
541	Db	 CGCACAGCAAGTCACCCGAACACCACTCTCAGTTTGT	TAGCAGAGCTGGATGAAGAAATAGT	600
1130	Qy	 TCTGCAAG	1137	
601	Db	 TCTGCAAG	608	

RESULT 9	770 bp	linear	EST 25-NOV-2002
BU204906			
LOCUS	604157793F1	CSEQCHN03	Gallus gallus cDNA clone CHEST1008j24 5',
DEFINITION			mRNA sequence.
ACCESSION	BU204906		
VERSION	BU204906.1	GI:25369490	
KEYWORDS			EST.
SOURCE			Gallus gallus (chicken)
ORGANISM			Gallus gallus
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
			Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
			Phasianinae; Gallus.
REFERENCE			1 (bases 1 to 770)
AUTHORS			Boardman, P.E., Sanz-Exquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
TITLE			A Comprehensive Collection of Chicken cDNAs
JOURNAL			Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE			22335534
PUBMED			12445392
CONTACT			Contact: Simon Hubbard
			Department of Biomolecular Sciences
			University of Manchester Institute of Science and Technology

FEATURES	
source	
/organism="Gallus gallus"	
/mol_type="mRNA"	
/strain="White Leghorn, Hisex"	
/db_xref="taxon:9031"	
/clone="ChEST1008724"	
/tissue_type="whole embryo"	
/dev_stage="20-21"	
Location/Qualifiers	
1. .770	
Email: simon.hubodargue@ucl.ac.uk.	
Db	598
GTGTGGACATTTGCAAGCAACATGCTGTAACTTCATCTTAGCAACACACAGCA	657
QY	1079
AGTACCCCG-AACACCAATTCACAGTTTGTATAACAGACGCTGGATGAAGAATAGCTCTGCAGG	1137
Db	658
AGTCCACCAAGAACACCATTCACAAATTTGTAGCAGAAGTGGATGAAGAATACCTTAG--GG	715
QY	1138
ACGAATGCCCTTAGT	1151
Db	716
ACGAATGCCCTTAGT	729

/lab\_host="DH10B"  
/clone\_lib="CSEQCHN03"  
/note="Organ: whole embryo; Vector: pBluescript II KS(+);  
Site 1: EcoRI; Site 2: NotI; This normalized library was  
constructed from 1 million independent clones. cDNA  
synthesis was initiated using an oligo(dT) primer, using  
methylated C in the first strand synthesis reaction.  
Following this first strand reaction, double-stranded cDNA  
was bluntend, ligated to NotI adapters, digested with  
EcoRI, size-selected, and cloned into the NotI and EcoRI  
compatible sites of a custom modified MCS of the  
pBluescript (KS+) vector. The library was normalized in 2  
rounds using conditions adapted from Soares et al., PNAS  
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
(1996): 791, except that a significantly longer  
reannealing hybridization was used."

## ORIGIN

Query Match	39.8%;	Score 512.8;	DB 5;	Length 770;
Best Local Similarity	83.9%;	Pred. No. 2.6e-129;		
Matches 616;	Conservative 0;	Mismatches 112;	Indels 6;	Gaps 3;
QY	419	TCAAAGTCTCTCCCAAAGCCACACCCCAAAGCCAGCTCTCTCTCCAGGGTCCCCCCCCA	478	
Db	1	TCAAAGTCTCTCCCAAAGCTACTACACCAAACCTGCGCGCATCTTCAAGGATCCCATCTA	60	
QY	479	TGAAGCTGGAATGGATGTGAGGCGTTTGGAGATGTCTCCAGTGAAGTCTCAACTTTGC	538	
Db	61	TGAAGTTGGAATAGATGTCGACGCTTTGAGGATGTCTCAACAGAAGTCTCCACTCTGC	120	
QY	539	ATATAAGAAAGCGCGGCGAGTCCAATCGAATCCTCAGCATCTTCTGTGATTTCTACAGCCC	598	
Db	121	ATATAAGAAAGGCGAGGCGAGTCAAACTGGAACCTCAGCATCTTCTGATTTCTGCAAGCTC	180	
QY	599	AGTTTGCCTCGAGCCTCTTCCAGACATCAGAGGGCAAACTACCTGTCTGTGATCTGGGCC	658	
Db	181	AGTGTCTTCCAGCCTCTTCCAGACATCTGAAGTAAATATTTATATCAGATCTAGGCC	240	
QY	659	CACAAGAGCGTATGCAAAATCTCTAAAGTTTACGGGACTCTCAATGACCACTATCAGTCACT	718	
Db	241	CACAGGAGCGTATGCAAGATCTCAAAATTTACTGGACTGTCAATGACCAACCATCAGCCACT	300	
QY	719	GGCTGCGCAACGTCGAAGTACCAAGCTTTAGGAAACGGCGGAGCAAAATTTCTGAAAAACA	778	
Db	301	GGTTGGCAATGTCAAGTACCAACTTAGGAAACTCGAGGAACAAGTTTTCGAAAAACA	360	
QY	779	TGGACAAAGGCCACCCCATCTTTTATTTGCAGTGACTGTGCTCCAGTTCAGAACCCCTT	838	
Db	361	TGGACAAAGGCCATCCGGTCTTTTATTTGCAGTGACTGTGCTATCTCAGTTTCGAACCCCAT	420	
QY	839	CTACCTACATCAGTCACTTAGAATCTCACTCGGTTTCCAAATGGAAGGACATGACCCGCT	898	
Db	421	CTACTTACATATGCCACTTAGAATCTCATCTAGGTTTCCAAATGGAAGACATGAACAGGC	480	
QY	899	TGTCAGTGGACCAAGCAAAAGCGTGAGCAAGAGATCTCCGGGTATCGTCGGCTCAGA	958	
Db	481	TGCGCTGTGAGCAGCAAAACCAAGGTAGCAAGAAATCTCCAGT---TTCAGTTCAA	537	
QY	959	GGTCTCCAGAAACAATAGCTGCGGAAGAGACAACAGATCTAAATTCAGGTGAAGTTGT	1018	
Db	538	GGTCTCCTGAAACAATAGCTGGAGAAGAGACAACAGATCTAAAGTTCAAAATGTAAGTGT	597	
QY	1019	GCTGTCCGACATTTGAGCAAAACATGCGGTAAACTCCACCTTAAGCAAAACCCACAGCA	1078	
Db	598	GCTGTCCGACATTTGCAAGCAAAACATGCTGTAAACCTTCATCTAAGCAAAACACACAGCA	657	
QY	1079	AGTCACCGC-AAACACCATTCACAGTTTGTAAACAGACGCTGGATGAAGAATAGCTCTGCAGG	1137	
Db	658	AGTCACAGAAACACCATTCACAAATTTGTAGCAGAAGTGGATGAAGAAATTAATTAG--GG	715	
QY	1138	ACGAATGCCCTAGT 1151		
Db	716	ACGAATGCCCTAGT 729		



RESULT 10	BF140337	953 bp	mRNA	linear	EST 24-OCT-2000			
LOCUS	601787846F1	NCI_CGAP_Lu30	Mus musculus	cDNA clone	IMAGE:4015643 5',			
DEFINITION	mRNA sequence.							
ACCESSION	BF140337							
VERSION	BF140337.1	GI:10979377						
KEYWORDS	EST.							
SOURCE	Mus musculus (house mouse)							
ORGANISM	Mus musculus							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.							
TITLE	1 (bases 1 to 953)							
JOURNAL	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .							
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps-f@mail.nih.gov">cgaps-f@mail.nih.gov</a> Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: NCI-CCAP Clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LHAM9261 row: 1 column: 12 High quality sequence stop: 676.							
FEATURES	Location/Qualifiers							
source	1..953							
	/organism="Mus musculus"							
	/mol_type="mRNA"							
	/strain="CZECH II"							
	/db_xref="taxon:10090"							
	/clone="IMAGE:4015643"							
	/tissue_type="tumor, metastatic to mammary"							
	/lab_host="DH10B"							
	/clone_lib="NCI CGAP Lu30"							
	/note="Organ: lung; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; transgenic model MNT-1, expression driven by MYTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"							
ORIGIN								
Query Match	39.2%; Score 505; DB 2; Length 953;							
Best Local Similarity	88.3%; Pred. No. 3.8e-127;							
Matches	628; Conservative 0; Mismatches 75; Indels 8; Gaps 7;							
Qy	418	GTCAAAGTCTCTCCCAAGGCCACCAACCCCAAGCCAGCCTCTCTCCAGGGTCCCCCCC 477						
Db	1	GTCAAGGTCTCTCCCAAGCCACCAACCCCAAGCCAGCGGCTCTCGAGGTCCCTCT 60						
Qy	478	ATGAAGCTGGAATGGATGTGAGCGCTTTGAGGATGTCTCCAGTGAAGTCTCAACTTG 537						
Db	61	ATGAAGCTGGAATCGAGCTTCAGACGCTTTGAGGATGTTTCGAGCGAAGTCTCCACTTG 120						
Qy	538	CATAAAA- GAAGAAGCGCGGAGTCCAACCTGAATCTCTCAGCATCTTCTGATTCTACAAGC 596						
Db	121	CACAAAAAGGAGAGAGCGCGGAGTCCAACCTGGAACCCCGAGCATCTTCTCATCTCTGCAAGC 180						
Qy	597	CCAGTTTGCCTCGAGCCTCTTCCAGACATCAGAGGGCAAAATACCTGTCTGTGATCTGGG 656						
Db	181	TCAG-TTGCTCTGAGCCTCTTCCAGACCTCAGAGGGCAAAATACCTGTCTGTGATCTGGG 239						
Qy	657	CCCAAGAGCGGTATGCAAAATCTCTTAAGTTTACGGGACTCTCTCAATGACCACATTCAGTCA 716						
Db	240	CCCCAAGAGCGGATGCAAAATCTCCAAGTTCACCGGACTCTCCATGACCACCAATCAGCCA 299						
Qy	717	CTGCTGCGCCAACTCAAGTACCAAGTTCAGAAACCGGGCGGACAAAATTTCTGAAAAA 776						
Db	300	CTGCTGCTAACCTCAAAATACCAAGTTCAGAAACCGGGCGGACAAAATTTCTGAAAAA 359						

the Not I and Eco RI sites of the modified pT7T3 vector.  
Library is normalized, and was constructed by Bento  
Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 38.3%; Score 493; DB 7; Length 494;  
Best Local Similarity 99.8%; Pred. No. 6.5e-124;  
Matches 493; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 299 AGAACAGCGATCAGCCGATTCAGCTCACCAGTCCAAAGGCAAGGCGAGTCTCTCCG 358  
Db 1 AGACAGCGATCAGCCGATTCAGCTCACCAGTCCAAAGGCAAGGCGAGTCTCTCCG 60  
QY 359 AAGCACAATCTTGTATGTCTCCACCTCAGAACGACGCTCTGTCTGACATCGCGGACATGG 418  
Db 61 AAGCACAATCTTGTATGTCTCCACCTCAGAACGACGCTCTGTCTGACATCGCGGACATGG 120  
QY 419 TCNAAGTCTCTCCCAAGGACACCCCAAGGCGGCTCTCTCCAGGGTCCCGCCCA 478  
Db 121 TCNAAGTCTCTCCCAAGGACACCCCAAGGCGGCTCTCTCCAGGGTCCCGCCCA 180  
QY 479 TGAAGCTGGAAATGGAATGAGTGTGAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTCG 538  
Db 181 TGAAGCTGGAAATGGAATGAGTGTGAGGCGCTTTGAGGATGCTCCAGTGAAGTCTCAACTTTCG 240  
QY 539 ATAAAGAAAAGCGCGGCGAGTCCAACTGGAATCCTCAGCATCTTCTGTATTTACAGGCC 598  
Db 241 ATAAAGAAAAGCGCGGCGAGTCCAACTGGAATCCTCAGCATCTTCTGTATTTACAGGCC 300  
QY 599 AGTTTCCTCGAGCCTCTTCCAGACATCAGAGGGGAAATACCTGCTGTCTGATCTGGGCC 658  
Db 301 AGTTTCCTCGAGCCTCTTCCAGACATCAGAGGGGAAATACCTGCTGTCTGATCTGGGCC 360  
QY 659 CACAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCATATCAGTCACT 718  
Db 361 CACAAGAGCGTATGCAAACTCTTAAGTTTACGGGACTCTCAATGACCATATCAGTCACT 420  
QY 719 GGCTGCCCAACGTCAGAGTACCAGCTTAGGAAACCGGGCGGACAAATTTCTGAAAAACA 778  
Db 421 GGCTGCCCAACGTCAGAGTACCAGCTTAGGAAACCGGGCGGACAAATTTCTGAAAAACA 480  
QY 779 TGGACAAAAGGCCAC 792  
Db 481 TGGACAAAAGGCCAC 494

RESULT 12  
AA873869/c  
LOCUS  
DEFINITION  
Ch79d06.s1 NCI\_CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:1473227 3',  
mRNA sequence.  
ACCESSION  
AA873869  
VERSION  
AA873869.1 GI:2968005  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 477)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 471.  
Location/Qualifiers  
1. 477  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1473227"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Kid3"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT7T3 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."

## ORIGIN

Query Match 36.9%; Score 475.4; DB 1; Length 477;  
Best Local Similarity 99.8%; Pred. No. 4.4e-119;  
Matches 476; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 789 CCACCCCATCTTTTATTGTCAGTGACTGTGCTCCCGAGTTCAGAACCCCTTCTACCTACAT 848  
Db 477 CCACCCCATCTTTTATTGTCAGTGACTGTGCTCCCGAGTTCAGAACCCCTTCTACCTACAT 418  
QY 849 CAGTCACCTAGAAATCTCACCTGGGTTTCAAAATGAAGACATGACCCGTTGTCTAGTGA 908  
Db 417 CAGTCACCTAGAAATCTCACCTGGGTTTCAAAATGAAGACATGACCCGTTGTCTAGTGA 358  
QY 909 CCAGCAAGAGAGGTGGAGAGAGATCTCCCGGGTATCGTCGGCTCAGAGTCTCCAGA 968  
Db 357 CCAGCAAGAGAGGTGGAGAGAGATCTCCCGGGTATCGTCGGTTCAGAGTCTCCAGA 298  
QY 969 AACAAATAGTGTCCGGAAGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTGTCTGGAC 1028  
Db 297 AACAAATAGTGTCCGGAAGAGGACACAGACTCTAAATTTCAAGTGTAAAGTTGTGTCTGGAC 238  
QY 1029 ATTTGTGAGCAAAATGCGGTAAACTCCACTAGCAAAAGCAGCAGCAAGTCAACCGA 1088  
Db 237 ATTTGTGAGCAAAATGCGGTAAACTCCACTAGCAAAAGCAGCAGCAAGTCAACCGA 178  
QY 1089 ACACCAATTCACAGTTTGTAAACAGAGCTGATCAAGAAATAGCTCTGCAGACCAATGCCCT 1148  
Db 177 ACACCAATTCACAGTTTGTAAACAGAGCTGATCAAGAAATAGCTCTGCAGACCAATGCCCT 118  
QY 1149 AGTTTCCACTTTCCAGCCTGGATCCCTCACAATGAACCCCTTCTTGTGTCACCAATCCTG 1208  
Db 117 AGTTTCCACTTTCCAGCCTGGATCCCTCACAATGAACCCCTTCTTGTGTCACCAATCCTG 58  
QY 1209 CTTCTGACATTTGAATCAATGAATCTCTCCAGACCCCTGGCTCTGAGAAAGTCTGCC 1265  
Db 57 CTTCTGACATTTGAATCAATGAATCTCTCCAGACCCCTGGCTCTGAGAAAGTCTGCC 1

## RESULT 13

CK773713/c  
LOCUS  
DEFINITION  
962388 MARC 2B0V Bos taurus cDNA 3', mRNA sequence.  
ACCESSION  
CK773713  
VERSION  
CK773713.1 GI:42727856  
KEYWORDS  
EST.  
SOURCE  
Bos taurus (cow)  
ORGANISM  
Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovinae; Bos.  
1 (bases 1 to 686)  
REFERENCE  
Smith, I.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,  
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C.,

Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Pertea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J., and Keele, J.W.  
 Sequence evaluation of four pooled-tissue normalized bovine cdna libraries and construction of a gene index for cattle  
 Genome Res. 11 (4), 626-630 (2001)  
 21180013  
 MEDLINE  
 COMMENT  
 Contact: Smith TPL  
 USDA, ARS, US Meat Animal Research Center  
 PO Box 166, Clay Center, NE 68933-0166, USA  
 Tel: 402 762 4366  
 Fax: 402 762 4390  
 Email: smithemail.marc.usda.gov  
 Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim\_alt option. Vector identified with cross match v0.990329.  
 Plate: 76 row: K column: 24  
 Seq primer: GTAATACGACTCACTATAGG.  
 Location/Qualifiers  
 1. .686  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_lib="MARC 2BOV"  
 /note="Vector: pcMV SPORT6; Site 1: Not1; Site 2: SalI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."

Query Match 36.1%; Score 465.2; DB 7; Length 686;  
 Best Local Similarity 88.2%; Pred No. 3e-116;  
 Matches 506; Conservative 0; Mismatches 68; Indels 0; Gaps 0;  
 444 CCCAAGCCAGCCTCTCTCCAGGGTCCCCCCCCATGAAGCTGGAAATGGATGTGAGCGG 503  
 |||||  
 686 CCCGAAGCCGCCACCTCCCCAGGGTCCCTCCCGTGAAGCTGGAGATGGATGTGAGCGG 627  
 |||||  
 504 CTTGAGATGTCTCCAGTGAAGTCTCACTTTGCATTAAGAAAGAGCGCGAGTCCAA 563  
 |||||  
 626 CTTGAGGACGTGTCCAGCGAGGTCTCCACCTTGATAGAGGAAAGCGCGAGTCCAA 567  
 |||||  
 564 CTGGAATCTCAGCATCTTGATTTCAAGCCAGTTTGCTCGAGCCTCTTCCAGAC 623  
 |||||  
 566 CTGGAATCCCCAGCATCTTGATCTCGAGGCTCAGTTTGCTCCAGCCTCTTCCAGAC 507  
 |||||  
 624 ATCAGAGGCAATACCTGTCTGTGATCTGGGCCCAAGAGCGTATGCAAAATCTCTAA 683  
 |||||  
 506 CTCGAGGGCAATACCTGTCTGTCCGACCTGGGCCCGCAAGAGCGGATGCGATCTCGAA 447  
 |||||  
 684 GTTTACGGGACTCTCAATGACCACTATCAGTCACTGTGGTGGCCAAAGTCAAGTACCACT 743  
 |||||  
 446 GTTTACGGGACTCTCGATGACCACTCAGCCTGTGGTGGCCAAAGTCAAGTACCACT 387  
 |||||  
 744 TAGGAAACGGCGGGACAAATTTCTGAAACATGACAAAGCCACCCCATCTTTTA 803  
 |||||  
 386 TAGGAAACGGCGGGACCAAGTTCTGAAACATGACAAAGCCACCCCATCTTTTA 327  
 |||||  
 804 TTGCAGTGACTGTGCTCTCCAGTTTCCAGAACCCCTTCTACCTACATCAGTCACTTAGAATC 863  
 |||||  
 326 CTGCAGTGACTGTGCTCTCCAGTTTCCGAACCCCTTCTACCTACATCAGCCTTAGAGTC 267  
 |||||  
 864 TCACCTGGGTTCCAAATGAAGGACATGACCCCGCTTGTGTCAGTGGACGAGCAAGCAAGGT 923  
 |||||  
 266 CCACCTGGGCTTCCAGATGAAGGACATGACCCCGCTGTGTCAGTGGACGAGCAAGGCAAGC 207  
 |||||  
 924 GGAGCAAGGATCTCCCGGTTATCTCGGCTCAGAGGCTCCAGAAACATAGTCTCCGA 983  
 |||||  
 206 GGAGCAGGAGATCTCCCGGTTGTCTCGGCTCAGCGGTCCCGGAAAGACAGTCTGGCGA 147  
 |||||

ORIGIN  
 source  
 1. .686  
 /organism="Bos taurus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9913"  
 /tissue\_type="pooled"  
 /lab\_host="DH10B"  
 /clone\_lib="MARC 2BOV"  
 /note="Vector: pcMV SPORT6; Site 1: Not1; Site 2: SalI; Library made from pooled tissue from testis, thymus, semitendinosus muscle, longissimus muscle, pancreas, adrenal, and endometrium."

Qy 984 AGAGGACACAGACTCTAAATTCAGTGAAGTTG 1017  
 |||||  
 Db 146 AGAGGACACAGGACTCTAAATTCAGTGAAGTTG 113  
 |||||  
 RESULT 14  
 LOCUS BO715061 1221 bp mRNA linear EST 16-JUL-2002  
 DEFINITION AGENCOURT 8474842 NIH\_MGC\_129 Mus musculus cdna clone IMAGE:6310221  
 5', mRNA sequence.  
 ACCESSION BO715061  
 VERSION BO715061.1 GI:21853960  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 1221)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgapbs@mail.nih.gov  
 Tissue Procurement: Susan L. Sullivan, Ph.D.  
 cDNA Library Preparation: Resgen, Invitrogen Corp  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LIAW1372 row: c column: 22  
 High quality sequence start: 79  
 High quality sequence stop: 516.  
 FEATURES  
 Location/Qualifiers  
 1. 1221  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:6310221"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 129"  
 /note="Organ: olfactory epithelium; Vector: pcMV-SPORT6.1;  
 Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally.  
 Primer: Oligo dt. Average insert size 2.2 kb. Constructed  
 by Resgen, Invitrogen Corp. Note: this is a NIH\_MGC  
 Library."  
 ORIGIN  
 Query Match 34.0%; Score 438; DB 5; Length 1221;  
 Best Local Similarity 88.0%; Pred. No. 1e-108;  
 Matches 500; Conservative 0; Mismatches 65; Indels 3; Gaps 2;  
 Qy 331 TCCAAAGCAGAGAGCGGAGTCTCGAAGCACAATCTTGTATGTCCTCCACCTCAGAG 390  
 |||||  
 Db 91 TCCAAAGCAGAGAGCGGAGTCTCGAAGCACAATCTTGTATGTCCTCCACCTCAGAG 150  
 |||||  
 Qy 391 CACGCTCTGTCTGACATCGCCGACATGTCCTCAAGTCTCTCCCAAGCCACCCCAAG 450  
 |||||  
 Db 151 CATGCTCTGTGATGATGTCGATGTCCTCCCAAGCCACCCCAAG 210  
 |||||  
 Qy 451 CCAGCTCTCTCTCAGGGTCCCCCATGAAGTGGAAATGGATGTCAGGGCTTTGAG 510  
 |||||  
 Db 211 CCAGCGGCTCTCAGGGTCCCTCTATGAAGTGGAAATCGAGCTCAGAGCTTCGAG 270  
 |||||  
 Qy 511 GATGCTCCAGTGAAGTCACTTTCATGAAGAGAGAGCGGAGTCCCACTGGAT 570  
 |||||  
 Db 271 GATGTTCCAGGAGAGTCTCCACTTTGCAAAAAGAGAGCGGAGTCCCACTGGAG 330  
 |||||  
 Qy 571 CCTCAGCATCTTCTGATTTCAAGCCAGTTTGTCTGAGCTCTTCCAGACATCAGAG 630  
 |||||  
 Db 331 CCCAGCATCTTCTCATCTCGAAGTCAAGTTTGTCTGAGCTCTTCCAGACCTCAGAG 390  
 |||||  
 Qy 631 GGCAATACCTGCTGTCTGATCTGGGCCCAAGAGCGTATGCAAAATCTCTAAGTTTACG 690  
 |||||

Db	391	GGCAAAATACCTCTTCTTGACCTGGGCCCCCAAGAGCGGATGCAAAATCTCAAGTTCCACC	450
Qy	691	GGAGCTCTCAATGACCACATATCAGTCACTGGCTGGCCCAACGTCAGAGTACCAGCTTTAGGAAA	750
Db	451	GGACTCTCATGACCAACATCAGCCACCTGGCTGGCTAACGTCACATACAGGCTTTAGGAAA	510
Qy	751	ACGGGGGGGACAAAATTTCTGAAAACATGGAACAAAGGCCACCCCATCTTTTATGTCAGT	810
Db	511	ACAGGTGGGACAAAGTTCTGAAAACATGACAAAGGGCACCCCATCTTTTACTGTCAGC	570
Qy	811	GACTGTGCTCCAGTTTCAGAA-CCCTTCTACCTACATCAGTCACTTAGATCTCACT	869
Db	571	GACTGTGCTCCAGTTTCACAAACCCCTCTACCTTAATCAGCCCTTAAAGTCTCCCT	630
Qy	870	GGG--TTTCCAAATGAAGACATGACCC	895
Db	631	GGGGCTTCCCAATGAAGACATGACCC	658
LOCUS	CK775937	581 bp	linear
DEFINITION	967078 MARC 380V Bos taurus cDNA 3', mRNA	EST 20-FEB-2004	
ACCESSION	CK775937		
VERSION	CK775937.1	GI:42732187	
KEYWORDS	EST.		
SOURCE	Bos taurus (cow)		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.		
AUTHORS	1 (bases 1 to 581) Smith, T.P.L., Grosje, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perle, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.		
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle		
JOURNAL	Genome Res. 11 (4), 626-630 (2001)		
MEDLINE	21180013		
PUBMED	11282978		
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called with phred v0.020425.c and trimmed with the aid of the trim_alt option. Vector identified with cross_match v0.990329. Plate: 119 row: P column: 16 Seq primer: GTAATACGACTCACTATAGG.		
FEATURES	Location/Qualifiers		
source	1..581 /organism="Bos taurus" /mol_type="mRNA" /db_xref="taxon:9913" /tissue_type="pooled" /lab_host="DH108" /clone_lib="MARC 380V" /note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI; Library made from pooled tissue from marrow, alveolar macrophage, ovary, fetal semitendinosus muscle, and fetal longissimus muscle."		
ORIGIN			
Query Match	33.8%;	Score 435.2;	DB 7; Length 581;
Best Local Similarity	86.1%;	Pred. No. Se-108;	
Matches 482;	Conservative 0;	Mismatches 78;	Indels 0; Gaps 0;
Qy	676	ATCTCTAAGTTTACGGGACTCTCAATGACCACTATCATGCTGCTGGCCAAAGCTCAAG	735

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:11:25 ; Search time 2220.75 Seconds  
(without alignments)  
9927.764 Million cell updates/sec

Title: US-08-731-499-8  
Perfect score: 455  
Sequence: 1 GAATCAGAGCTTTAATATG.....ATTGCTCTATTTCNGGGGGT 455

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues  
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.on.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.ats.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	453	99.6	455	6	AR070331	Sequence
2	453	99.6	455	6	BD085732	Genes fro
3	428	94.1	4723	9	BC071570	Homo sapi
4	428	94.1	4944	9	AB002385	Human mRN
5	428	94.1	118947	2	AC006321	Homo sapi
6	426.4	93.7	2328	6	AR369583	Sequence
7	426.4	93.7	2328	6	AR404162	Sequence
8	426	93.6	4699	9	HSU81561	Human prote
9	424.8	93.4	190846	9	AC006372	Homo sapi
10	416	91.4	4735	9	BC034040	Homo sapi
11	414.6	91.1	4719	9	HSU66702	Human phogr
12	105	23.1	105	11	GI9707	human STS A
13	86.2	18.9	180187	2	AC117965	Rattus no
14	85.6	18.8	5136	10	RNU73458	Rattus norv
15	76.8	16.9	222251	2	AC110936	Rattus no
16	74.6	16.4	197934	2	AC117676	Mus muscu
17	74.6	16.4	211959	2	AC118023	Mus muscu
18	50.4	11.1	105604	9	AC084013	Homo sapi
19	49.8	10.9	106216	9	AC061997	Homo sapi

20	49.8	10.9	255522	2	AC129169	Rattus no
21	49.4	10.9	51	6	CQ003853	Sequence
22	49	10.8	249943	3	AE014823	Plasmodi
23	48	10.5	75076	9	AC004948	Homo sapi
24	47.8	10.5	103344	9	HS1100E15	Human DNA
25	46.6	10.2	82139	3	AC115684	Dicystoate
26	46.6	10.2	149591	2	AC141821	Apis mell
27	46.4	10.2	191481	2	AC108123	Homo sapi
28	46	10.1	144759	9	BS352A20	Human DNA
29	46	10.1	162905	2	EX927388	Danio rer
30	45.8	10.1	127230	9	AC027296	Homo sapi
31	45.8	10.1	173997	2	AC012340	Homo sapi
32	45.8	10.1	176898	9	AC117569	Homo sapi
33	45.8	10.1	194028	2	AC111077	Homo sapi
34	45.8	10.1	229569	2	AC107424	Homo sapi
35	45.6	10.0	158548	3	PFMAL3P2	Plasmodi
36	45.6	10.0	272384	2	AC125722	Rattus no
37	45.4	10.0	2000	6	AX655393	Sequence
38	45.4	10.0	150754	9	AC023491	Homo sapi
39	45.4	10.0	132884	9	AC138625	Homo sapi
40	45.4	10.0	156168	2	EX927212	Danio rer
41	45.4	10.0	158420	9	AC137788	Homo sapi
42	45.4	10.0	175358	9	AC007981	Homo sapi
43	45.4	10.0	192049	2	AC137785	Homo sapi
44	45.2	9.9	770	11	BV006800	STS_CBB26
45	45.2	9.9	8099	9	AP001114	Homo sapi

ALIGNMENTS

RESULT 1  
AR070331  
LOCUS  
DEFINITION Sequence 8 from patent US 5892010.  
ACCESSION AR070331  
VERSION AR070331.1 GI:7221219  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 455)  
AUTHORS Gray,J., Collins,C., Hwang,S.-i., Godfrey,T., Kowbel,D. and Rommens,J.  
TITLE Genes from the 20Q13 amplicon and their uses  
JOURNAL Patent: US 5892010-A 8 06-APR-1999;  
FEATURES Location/Qualifiers  
source 1..455  
/organism="unknown"  
/mol\_type="unassigned DNA"

ORIGIN	Query Match	99.6%	Score 453;	DB 6;	Length 455;
	Best Local Similarity	100.0%	Pred. No. 6.2e-102;	Mismatches 0;	Indels 0; Gaps 0;
	Matches	455;	Conservative 0;		
Qy	1	GAATCAGAGCTTTAATATGACACAAATTAATATATTTTGTATATCTCACC	GGAGNTTC	60	
Db	1	GAATCAGAGCTTTAATATGACACAAATTAATATATTTTGTATATCTCACC	GGAGNTTC	60	
Qy	61	TCCTCAACATAGAGGTTAGAAATTAACAGTAGGCATATGCTTCCTATATTCAGATAA	120		
Db	61	TCCTCAACATAGAGGTTAGAAATTAACAGTAGGCATATGCTTCCTATATTCAGATAA	120		
Qy	121	TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAG	180		
Db	121	TTCAATTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAG	180		
Qy	181	CTATATTAAGCAGATATTCATTACAATACCATGTAGAGACATAGCAATATTTTGGCAT	240		
Db	181	CTATATTAAGCAGATATTCATTACAATACCATGTAGAGACATAGCAATATTTTGGCAT	240		
Qy	241	CATTCTGTCCGCTCAGTAGAGGGCGTGTTCCCTCTGTGGAGAGTACCATCTA	300		

Db	241	TCTGTCAGTAGGCGCGTGTTCCTCTGTTAGGCGCTTTTGAGAGTACCATCTA	300
Qy	301	TCTAAGATGGAGAAATCTGTGGGAGGCGGGATGGAGGTGGTTCACGCTGAACC	360
Db	301	TCTAAGATGGAGAAATCTGTGTGGGAGGCGGGATGGAGGTGGTTCACGCTGAACC	360
Qy	361	CCACACAGGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGCGCTTCCATGTGATCATCC	420
Db	361	CCACACAGGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGCGCTTCCATGTGATCATCC	420
Qy	421	TGCTCAATGAAGTGAATTCCTATTTCNGGGGGT	455
Db	421	TGCTCAATGAAGTGAATTCCTATTTCNGGGGGT	455
RESULT 2			
BD085732			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Genes from the 20q13 amplicon and their uses.			
BD085732			
BD085732.1			
JP 2001524802-A/8.			
synthetic construct			
other sequences; artificial sequences.			
1 (bases 1 to 455)			
Gray,J.W., Collins,C.C., Hwang,S.I., Godfrey,T., Kowbel,D. and			
Rommens,J.			
Genes from the 20q13 amplicon and their uses			
Patent: JP 2001524802-A 8 04-DEC-2001;			
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA			
OS Artificial Sequence			
PN JP 2001524802-A/8			
FD 04-DEC-2001			
PF 15-JUL-1997 JP 1998506264			
PR 15-JUL-1996 US 08/680395,16-OCT-1996 US 08/731499 PR			
17-JAN-1997 US 08/785532			
PI JOE W GRAY, COLIN CONRAD COLLINS, SOO IN HWANG, TONY GODFREY, PI			
DAVID KOWBEL,			
PI JOHANNA ROMMENS			
PC C12N15/11.C12Q1/68.A61K48/00			
CC Description of Artificial Sequence:20sa7			
CC n is A, C, G, T, or U			
FH Key Location/Qualifiers			
FT modified base (57)			
FT modified base (449).			
FEATURES			
source			
1. .455			
/organism="synthetic construct"			
/mol_type="genomic DNA"			
/db_xref="taxon:32630"			
ORIGIN			
Query Match			
Best Local Similarity 99.6%; Score 453; DB 6; Length 455;			
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	GAAATCAGAAGTTTAATATGACACAATTAATATTTGTAATCTCACCACCGAGNTTC	60
Db	1	GAAATCAGAAGTTTAATATGACACAATTAATATTTGTAATCTCACCACCGAGNTTC	60
Qy	61	TCTTCAACATAAGGAGTTAGAAATACAGTAGGCATATGCTTCCTATATTCAGATAAA	120
Db	61	TCTTCAACATAAGGAGTTAGAAATACAGTAGGCATATGCTTCCTATATTCAGATAAA	120
Qy	121	TTCAATTCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAGAAATGATAG	180
Db	121	TTCAATTCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAGAAATGATAG	180
Qy	181	CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACGAATATTTTGGCAT	240
Db	181	CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACGAATATTTTGGCAT	240

Qy	241	CATTCTGTCGCTCAGTAGGCGCGTGTTCCTCTGTTAGGCGCTTTTGAGAGTACCATCTA	300
Db	241	CATTCTGTCGCTCAGTAGGCGCGTGTTCCTCTGTTAGGCGCTTTTGAGAGTACCATCTA	300
Qy	301	TCTAAGATGGAGAAATCTGTTCGGGAGGCGGGATGGAGGTGGTTCACGCTGAACC	360
Db	301	TCTAAGATGGAGAAATCTGTTCGGGAGGCGGGATGGAGGTGGTTCACGCTGAACC	360
Qy	361	CCACACAGGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGCGCTTCCATGTGATCATCC	420
Db	361	CCACACAGGAAATCTGCAGGCCACACAGCTGCCTCTGCGCGCGCTTCCATGTGATCATCC	420
Qy	421	TGCTCAATGAAGTGAATTCCTATTTCNGGGGGT	455
Db	421	TGCTCAATGAAGTGAATTCCTATTTCNGGGGGT	455
RESULT 3			
BC071570/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
1 (bases 1 to 4723)			
Strausberg R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,			
Klausner,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D.,			
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,			
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,			
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,			
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,			
Schetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,			
Carninci,P., Frange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,			
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,			
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,			
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,			
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,			
Fahy,J., Helton,E., Kettman,M., Madan,A., Young,A.C., Rodrigues,S.,			
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,			
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,			
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.B.,			
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.			
Generation and initial analysis of more than 15,000 full-length			
human and mouse cDNA sequences			
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)			
12477932			
2 (bases 1 to 4723)			
Strausberg,R.			
Direct Submission			
Submitted (01-JUN-2004) National Institutes of Health, Mammalian			
Gene Collection (MGC), Cancer Genomics Office, National Cancer			
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,			
USA			
NIH-MGC Project URL: <a href="http://mgc.ncbi.nlm.nih.gov">http://mgc.ncbi.nlm.nih.gov</a>			
Contact: MGC help desk			
Email: <a href="mailto:cgabs-r@mail.nih.gov">cgabs-r@mail.nih.gov</a>			
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.			
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki			
Toshiyuki and Piero Carninci (RIKEN)			
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)			
DNA Sequencing by: Sequencing Group at the Stanford Human Genome			
Center, Stanford University School of Medicine, Stanford, CA 94305			
Web site: <a href="http://www.shgc.stanford.edu">http://www.shgc.stanford.edu</a>			
Contact: (Dickson, Mark) <a href="mailto:mcd@paxil.stanford.edu">mcd@paxil.stanford.edu</a>			
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,			

R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 167 Row: b Column: 15  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 19743910.

## FEATURES

## source

1. 4723  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:87154 IMAGE:4375815"  
/tissue\_type="Brain, hippocampus"  
/clone\_lib="NIH MGC\_95"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"

## gene

1. 4723  
/gene="PTPRN2"  
/note="synonyms: IA-2beta, IAR, ICAR, KIAA0387, PTPRP, IAR PTPRP"  
/db\_xref="LocusID:5799"  
/db\_xref="MIM:601698"

## CDS

120..2021  
/gene="PTPRN2"  
/codon\_start=1  
/product="PTPRN2 protein"  
/protein\_id="AAH71570.1"  
/db\_xref="GI:47939489"  
/db\_xref="LocusID:5799"  
/db\_xref="MIM:601698"  
/translation="MGPLPLLLLLPRVLPAAPSSVPRGRQLPGRIGCLLEEG  
LCGASEACVNGFTWDDYTQVMDQELADLPKTYLRPEASSPARPKSHVSGSRRYS  
REGGAALANLRULPFLPEALSPASDLARTAOQPPAEGDPRFSEILTYVAH  
TSALTYPPQPTOLREDLLPETLQLOPDELSPKVDGVDHRLHMAALSAVAOQPPA  
PPQGSLEPQVLLRAPSMPRLPAAPQKWPPLGSDSPSTGSGARHTLTKOL  
RQPAEVRVLSGLELDGMAELMAGLMQGVHGVARGSPRAALGESGEQADGPKTLR  
GDSFPDDGVQDDRLQYOEVRHLSATLGLLQDHGSRLLPGALTFAPLDMERKKSEH  
PESLSSEETAGDENVKVSQYKDLLQCPHSPGAAAFGELOQMPGPKSQSLP  
AGAEALSDGLQLEVPQSEEARGYIVTDRLPAPEGRRLVEDVARLLQVPSAFAD  
VEVLGPATVFKVSANQVNTEDVEKATVDNKKLEETSGLKILQTVGSGSKLFLP  
FOAQEDSTKFIALTLSLACILGVLLASGLIYCLRHSSQHRLEKSLGGLGDPGADA  
TAAYQELCR"

## ORIGIN

Query Match 94.1%; Score 428; DB 9; Length 4723;  
Best Local Similarity 99.1%; Pred. No. 1.2e-95;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNNT 59  
DB 4705 GAAATCAGAAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGGAGTTT 4646

QY 60 CTCCTTCAACATAAGAGGTAGAAATTAACAGTAGGCATATGCTTCCTATATTCAGATAA 119  
DB 4645 CTCCTTCAACATAAGAGGTAGAAATTAACAGTAGGCATATGCTTCCTATATTCAGATAA 4586

QY 120 ATTCATTTTCGATTAAATTAATCCAGATAGAGAGTAATTTTCGGAAGAAAGATGATA 179  
DB 4585 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGTAATTTTCGGAAGAAAGATGATA 4527

QY 180 GCTATATTAAAGCAGATATTCATTACATACCATGTAGACATAGCAATATTTTGGCA 239  
DB 4526 GCTATATTAAAGCAGATATTCATTACATACCATGTAGACATAGCAATATTTTGGCA 4467

QY 240 TCATTTCTGCTCGCTCAGTAGGCGGTGTTCCCTCTCGTAGGCGCTTTGGAGAGTACCATCT 299  
DB 4466 TCATTTCTGCTCGCTCAGTAGGCGGTGTTCCCTCTCGTAGGCGCTTTGGAGAGTACCATCT 4407

QY 300 ATCTTAAGATGGAGGAATGCTGTGGGAAGCGGGATGAGAGTGCGTTTTCTACGCTGAAC 359  
DB 4406 ATCTTAAGATGGAGGAATGCTGTGGGAAGCGGGATGAGAGTGCGTTTTCTACGCTGAAC 4347

```
SDAGRSCTHYVLIDVNLKMAKGAKEIDIAATLEHLRDORPGMVQTKQEFEPALTAVA
EEVNAITLKALPQ"

Query Match          94.1%; Score 428; DB 9; Length 4944;
Best Local Similarity 99.1%; Pred. No. 1.2e-95;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAATATGACACAAATTAATAATATTTGTATATCTTCACACCGGA-GN1T 59
Db 4942 GAAATCAGAAGTTTAATATGACACAAATTAATAATATTTGTATATCTTCACACCGGAGTTT 4883

QY 60 CTCCTCAACATAGAGGTAGAAATTAACAAGTAGCATATGCTTCTCTATATTCAGATAA 119
Db 4882 CTCCTCAACATAGAGGTAGAAATTAACAAGTAGCATATGCTTCTCTATATTCAGATAA 4823

QY 120 ATTCATTTTCGATTAATTAATTCACATAGAGAGTAATTTTCGAAAAAGAAATGATA 179
Db 4822 ATTCATTTTCGATTAATTAATTCACATAGAGAGTAATTTT-GAAAAAGAAATGATA 4764

QY 180 GCTATATTAAGCAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 4763 GCTATATTAAGCAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCA 4704

QY 240 TCATTCTGTCGCTCAGTAGGCGTGTTCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299
Db 4703 TCATTCTGTCGCTCAGTAGGCGTGTTCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 4644

QY 300 ATCTAAGATGGAGGATGCTGTGGGAGGGCGGATGGAGGTGCGTTTCTACGCTGAAC 359
Db 4643 ATCTAAGATGGAGGATGCTGTGGGAGGGCGGATGGAGGTGCGTTTCTACGCTGAAC 4584

QY 360 CCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 419
Db 4583 CCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 4524

QY 420 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 453
Db 4523 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 4490

RESULT 5
AC006321 118947 bp DNA linear HTG 24-AUG-2000
LOCUS Homo sapiens chromosome UNK clone RP5-991P5, WORKING DRAFT
DEFINITION SEQUENCE, 2 unordered pieces.
ACCESSION AC006321
VERSION AC006321.3 GI:9887769
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118947)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 118947)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 24, 2000 this sequence version replaced gi:8018216.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- project information -----
Center project name: H_DJ0991P05
----- Summary Statistics -----
Sequencing vector: M13; 85%
```

```
Sequencing vector: plasmid; 15%
Chemistry: Dye-primer ET; 85% of reads
Chemistry: Dye-terminator Big Dye; 15% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 117806 bases at least Q40
Consensus quality: 118306 bases at least Q30
Consensus quality: 118463 bases at least Q20
Insert size: 132000; agarose-fp
Insert size: 118847; sum-of-contigs
Quality coverage: in Q20 bases; agarose-fp
Quality coverage: in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 92854: contig of 92854 bp in length
* 92855 92854: gap of unknown length
* 92955 118947: contig of 25993 bp in length.
FEATURES
    source
        1..118947
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="UNK"
            /clone="RP5-991P5"
            1..92854
            /note="assembly_name:Contig10
            clone end:SP6
            vector side:right"
            92955..118947
            /note="assembly_name:Contig9
            clone end:T7
            vector side:left"
        misc_feature
            92854
        misc_feature
            92955
ORIGIN
Query Match          94.1%; Score 428; DB 2; Length 118947;
Best Local Similarity 99.1%; Pred. No. 1.4e-95;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAATATGACACAAATTAATAATATTTGTATATCTTCACACCGGA-GN1T 59
Db 29867 GAAATCAGAAGTTTAATATGACACAAATTAATAATATTTGTATATCTTCACACCGGAGTTT 29926

QY 60 CTCCTCAACATAGAGGTAGAAATTAACAAGTAGCATATGCTTCTCTATATTCAGATAA 119
Db 29927 CTCCTCAACATAGAGGTAGAAATTAACAAGTAGCATATGCTTCTCTATATTCAGATAA 29986

QY 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAAGAAATGATA 179
Db 29987 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTT-GAAAAAGAAATGATA 30045

QY 180 GCTATATTAAGCAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 30046 GCTATATTAAGCAGATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCA 30105

QY 240 TCATTCTGTCGCTCAGTAGGCGGTTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299
Db 30106 TCATTCTGTCGCTCAGTAGGCGGTTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 30165

QY 300 ATCTAAGATGGAGGATGCTGTGGGAGGGCGGATGGAGGTGCGTTTCTACGCTGAAC 359
Db 30166 ATCTAAGATGGAGGATGCTGTGGGAGGGCGGATGGAGGTGCGTTTCTACGCTGAAC 30225

QY 360 CCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 419
Db 30226 CCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 30285

QY 420 CTGGTCAATGAAGTGAATTTGCTATTTCNCGGG 453
```



1 (bases 1 to 2328)



## COMMENT

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On May 5, 1999 this sequence version replaced gi:4139346.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0331D05  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

## MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

## SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tacheno, M., Catanese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACE3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-703P8. Actual start of this clone is at base position 1 of RP11-331D5 actual end is at 190846 of RP11-331D5.

## FEATURES

```
source
    location/Qualifiers
    1..190846
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="7"
    /map="7"
    /clone="RP11-331D5"
    /clone_lib="RPCI-11"
    342..476
    /rpt_family="L1"
    1730..1793
    /rpt_family="L2"
    2543..2854
    /rpt_family="Alu"
    3541..3836
    /rpt_family="Alu"
    4686..4995
    /rpt_family="Alu"
    5406..5572
    /rpt_family="L1"
    6105..6258
    /rpt_family="Alu"
    6534..6809
    /rpt_family="Alu"

repeat_region
    7392..7661
    /rpt_family="Alu"

repeat_region
    7667..7967
    /rpt_family="Alu"

repeat_region
    9590..9745
    /rpt_family="Alu"

repeat_region
    11186..11316
    /rpt_family="MER2_type"

repeat_region
    11395..11687
    /rpt_family="Alu"

misc_feature
    11946..12180
    /note="CpG island (%GC=63.8, o/e=0.80, #CpGs=20)"

repeat_region
    12930..13229
    /rpt_family="Alu"

repeat_region
    13846..13969
    /rpt_family="Alu"

repeat_region
    13976..14020
    /rpt_family="MIR"

repeat_region
    14625..15183
    /rpt_family="ERV1"

repeat_region
    15184..15355
    /rpt_family="Alu"

repeat_region
    15356..15471
    /rpt_family="ERV1"

repeat_region
    15699..15992
    /rpt_family="Alu"

repeat_region
    16004..16221
    /rpt_family="Alu"

repeat_region
    16222..16524
    /rpt_family="Alu"

repeat_region
    16525..16652
    /rpt_family="Alu"

misc_feature
    21396..22150
    /note="CpG island (%GC=68.7, o/e=0.74, #CpGs=72)"

gene
    21514..29158
    /gene="DNAJB6"

mRNA
    join(21514..21720,27735..29158)
    /gene="DNAJB6"

CDS
    join(<21514..21720,27735..27817)
    /gene="DNAJB6"
    /note="Homo sapiens, MRJ gene for a member of the DNAJ protein family, clone MGC:1152 IMAGE:3346070, mRNA, complete cds.; H_NH0331D05.1"
    This gene was based on gi(12803262)
    Continues from H_T02506A20.1.2"
    /codon_start=3
    /product="unknown"
    /protein_id="AA07521.1"
    /db_xref="GI:41473761"

misc_feature
    23374..23612
    /gene="DNAJB6"
    /note="CpG island (%GC=66.5, o/e=0.69, #CpGs=18)"

repeat_region
    23786..23910
    /rpt_family="ERV1"

repeat_region
    25438..25739
    /rpt_family="Alu"

misc_feature
    25960..26373
    /gene="DNAJB6"
    /note="CpG island (%GC=65.0, o/e=0.73, #CpGs=33)"

Query Match
    93.4%; Score 424.8; DB 9; Length 190846;
Best Local Similarity
    98.7%; Pred. No. 9.2e-95;
Matches 448; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

Qy 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATTGTTATCTACACCGGA-GWTT 59
Db 150778 GAAATCAGAAGTTTAAATATGACACAAATTAATATTGTTATCTACACCGGAGTTT 150837

Qy 60 CTCTTCAACATAGGAGTTAGAAATTCACAGTAGGCGATGCTTCTATATTCAGATAA 119
Db 150838 CTCTTCAACATAGGAGTTAGAAATTCACAGTAGGCGATGCTTCTATATTCAGATAA 150897
```

QY 120 ATTCTTTTCGATTAAATTAATCCAGATAGAGAGAGTAATTTTCGAAAGAAATGATA 179  
 |||||  
 DB 150898 ATTCTTTTCGATTAAATTAATCCAGATAGAGAGAGTAATTTT-GGAAAGAAATGATA 150956  
 |||||  
 QY 180 GCTATATTAAACAGATATTCATTCAATACCATGTAGACACATAGCAATATTTTGCCA 239  
 |||||  
 DB 150957 GCTATATTAAACAGATATTCATTCAATACCATGTAGACACATAGCAATATTTTGCCA 151016  
 |||||  
 QY 240 TCATTCTGTCCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGATACATCT 299  
 |||||  
 DB 151017 TCATTCTGTCCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGATACATCT 151076  
 |||||  
 QY 300 ATCTAAGTGGAGGAATGCTGTGGGAAGCGCGGATGGAGTGCCTTTCTACGCTGAAC 359  
 |||||  
 DB 151077 ATCTAAGTGGAGGAATGCTGTGGGAAGCGCGGATGGAGTGCCTTTCTACGCTGAAC 151136  
 |||||  
 QY 360 CCCACACAGGAATCTGCAGCCACACACAGCTGCCTCTGCGCGCTTCATGTATCATC 419  
 |||||  
 DB 151137 CCCACACAGGAATCTGCAGCCACACACAGCTGCCTCTGCGCGCTTCATGTATCATC 151196  
 |||||  
 QY 420 CTGGTCATGAGTGAATGCTGCTATTTTCNGGGG 453  
 |||||  
 DB 151197 TTGGTCAATGAATGAATGCTCTATTTCTGGGG 151230  
 |||||

RESULT 10  
 BC034040/c  
 LOCUS  
 DEFINITION Homo sapiens protein tyrosine phosphatase, receptor type, N polypeptide 2, transcript variant 3, mRNA (cDNA clone MGC:26506 IMAGE:4811578), complete cds.  
 ACCESSION BC034040  
 VERSION BC034040.1 GI:21707838  
 KEYWORDS  
 SOURCE MGC.  
 ORGANISM Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 4735)  
 Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,B., Moore,T., Max,S.I., Wang,J., Hsieh,P., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.B., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smalls,D.E., Scherch,A., Schein,J.E., Jones,S.J. and Marra,M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
 12477932  
 2 (bases 1 to 4735)  
 Strausberg,R.  
 Direct Submission  
 Submitted (02-JUL-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: http://mgi.nci.nih.gov  
 Contact: MGC help desk  
 Email: cgapps-r@mail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: http://www-shgc.stanford.edu  
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: IRAK Plate: 32 Row: i Column: 20  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 19743910.

# FEATURES

## source

1..4735  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:26506 IMAGE:4811578"  
 /tissue\_type="Brain, hippocampus"  
 /clone\_lib="NIH\_MGC\_95"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"  
 1..4735  
 /gene="PTPRN2"  
 /note="synonyms: IA-2beta, IAR, ICAAR, KIAA0387, PTPRP, IAR PTPRP"  
 /db\_xref="LocusID:5799"  
 /db\_xref="MIM:601698"  
 105..3065  
 /gene="PTPRN2"  
 /codon\_start=1  
 /product="protein tyrosine phosphatase, receptor type, N polypeptide 2, isoform 3 precursor"  
 /protein\_id="AAH34040.1"  
 /db\_xref="GI:21707839"  
 /db\_xref="LocusID:5799"  
 /db\_xref="MIM:601698"  
 /translation="MGPEPLPPLLLLLLPPRLPAAPSSVPRQLPGRGLCLLEEG LCGSEACVNDGVFCGRQKPMDFYEVSPVALQRKVALQKLSGTGFTQDDYITQ YMQDELADLPKLYLRPEASPARPSKHSVSGRRYREGGAALANALRRHLPFLLEA LSQAPASDLARTHTAQRPAEGDDRFSESLTYVAHTSALTYPGPTQLREDDLLP RLAPAPAKWPSPLGSDPSSTGDGARHTLLKDLQROPAEYVGLSLDGMAE LMAGLMQGVHGVARGSPRALGSGSQADGPKATLGDSPFDGVDODDDRLYOEV HRLSATLGLLQDHGSRLLPGLPAPFAPLDMERKKSHPESLSSEETAGVENVKSQ TYSKLLGQHPSEFGAAAGFELQNMFGPKSEQLPAGAQLSDGLQLEVPSEE EARGYIVTDREVLGPVTFKVSANQVNTDEVATVDNKDLSETSGKILQTVG SKSKLFLPPQAEQSDTKFTALTJLVSLACILGVLASGLIYCLRHSQRLKEKLSG LGSDPGADATAYQELCRQRMATRPDRPEGPTSRISVSSQSDGPTSPSARSSA SWSSEFPQSNMDISTGHMILSYMEDHLKNKLEKEALCAYQAEPNSSFAQREE NVPKNRSLAVLYDHSRVLKKAENSHSDYINASPIHMDPRNPAYIATQGPLPATV LVRFWLMQVWESGCVIWMLTPLAENGVRQCYHYWPDGNSLHYIYEVNLVSEHCBDF IDRSYLNKQLTNETRTVTFQHFSLGWYDRGVSPSSRLDFFRKYNKCYGRSCPTIV HCSDCAGRSQTVLIDMVLNKMKAKEIDIAATLEHLRDRQRPQGVQTEQEFALTA VAEVYNAILKALPQ"

## gene

## CDS

## ORIGIN

Query Match 91.4%; Score 416; DB 9; Length 4735;  
 Best Local Similarity 98.9%; Pred. No. 1,1e-92;  
 Matches 449; Conservative 0; Mismatches 2; Indels 3; Gaps 3;  
 QY 1 GAAATCAGAGTTTAAATATGACACAAATTAATATTTTGTATATCTCACACCGGA-GNTT 59  
 |||||  
 DB 4716 GAAATCAGAGTTTAAATATGACACAAATTAATATTTTGTATATCTCACACCGAGGTTT 4657  
 |||||  
 QY 60 CTCTTCAACATAGAGGAGTTAGAAATTACAGATAGAGAGTAAATCTCTATATTCAGATAA 119  
 |||||  
 DB 4656 CTCTTCAACATAGAGGAGTTAGAAATTACAGATAGAGATATGCTTCTATATTCAGATAA 4597  
 |||||  
 QY 120 ATTCAATTCGATTAATAATTAATTCAGATAGAGAGAGTAAATTTTCGGAAGAAAGATGATA 179  
 |||||

```
|||||
4586 ATTCAATTCGATTAAATAATCCAGATAGAGAGAAATATTTT-GGAAAAGAAATGATA 4538
Db
180 GCTATATTAAAGCAGATATTCAATTAACAATACCATGTAGACATAAGCAATATTTTGGCA 239
Qy
4537 GCTATATTAAAGCAGATATTCAATTAACAATACCATGTAGACATAAGCAATATTTTGGCA 4478
Db
240 TCATTCCTCCGCTCAGTAGCGCGTGGTCCCTCTGGTAGAGCGCTTTGGAGAGTACCACTCT 299
Qy
4477 TCATTCCTCCGCTCAGTAGCGCGTGGTCCCTCTGGTAGAGCGCTTT-GAGAGTACCACTCT 4419
Db
300 ATCTAAGATGAGGAATCTGTGGAGCGGATGCGATGCGTCTTCTACGCTGAAC 359
Qy
4418 ATCTAAGATGAGGAATCTGTGGAGCGGATGCGATGCGTCTTCTACGCTGAAC 4359
Db
360 CCCACACAGGAATCTGCAGGCCACACACAGCTGCTCTGCGCGCTTCCATCTGATCATC 419
Qy
4358 CCCACACAGGAATCTGCAGGCCACACACAGCTGCTCTGCGCGCTTCCATCTGATCATC 4299
Db
420 CTGGTCAATGAAGTGAATTCCTATTTTCNGGGG 453
Qy
4298 CTGGTCAATGAAGTGAATTCCTATTTCTGGG 4265
Db

RESULT 11
HSU66702/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 4719)
AUTHORS
Kawasaki, E., Hutton, J.C. and Eisenbarth, G.S.
TITLE
Molecular cloning and characterization of the human transmembrane
protein tyrosine phosphatase homologue, phogrin, an autoantigen of
type 1 diabetes
JOURNAL
Biochem. Biophys. Res. Commun. 227 (2), 440-447 (1996)
MEDLINE
97032784
PUBMED
8878534
REFERENCE
2 (bases 1 to 4719)
AUTHORS
Kawasaki, E., Hutton, J.C. and Eisenbarth, G.S.
TITLE
Direct Submision
JOURNAL
Submitted (12-AUG-1996) Barbara Davis Center for Childhood
Diabetes, University of Colorado Health Sciences Center, 4200 East
9th Avenue, Box B-140, Denver, CO 80262, USA
FEATURES
Location/Qualifiers
1..4719
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="islet"
/tissue_type="pancreas"
15..3052
/note="protein tyrosine phosphatase homolog"
/codon_start=1
/product="phogrin"
/db_xref="GI:1620664"
/translation="MGPPPLPILLLLLLPPRVLPAPSPVPRGRLPGLGCLLEEG
LCGASEACVNDGVRCOKVPAMDYRVEVSVVALQRLVALOKLGGFTWDDYTO
YVMDQLADLPKTYLRRPEASPARSKHSVGSERYREGGAALANALRHLPLEA
LSQAPASDLAKTHQDRPPAGDDRFSESLITYVAHTSAUTYPGRTQREDLLP
RTLQGLQDELSPKVDSVDGRHLMALSAVAQPPAPPGEGSLEPOYLRLAPSRMP
RPLAPAPQKWPSPSGSDPSSTGDGARITHLLKDLQRPFAEVRGLSGLEADGMAE
LMAGLMQGVHDGVAGSPGALGEGADGPKATLRGDFPDGVDODDRLYQEV
HRLSATLGLLQDHGSRLLPGALPEARPLDMERKKSEHPSSLSSEETAGVENVKSO
TYSKDLGQQHPSEGAFAFELQNMFGPKSEQSLPAGAEALSDGLQLEVQSESE
EARGYIVTRDLPREGRRLVEDVARLLQVPSSAFADVEVLGPVAVTFKVSANQVNT
TEDVEKATVDNKKDLEETSGLKILQTVGSGSKLKLFLPPQAEQEDSTKFIALTLSLA
```

## ORIGIN

```
Query Match      91.1%; Score 414.6; DB 9; Length 4719;
Best Local Similarity 98.2%; Pred. No. 2.5e-92;
Matches 439; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Qy 8 GAAAGTTTAATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNTTCTCTTCA 66
Db 4710 GAAAGTTTAATGACACAAATTAATATATTTGTATATCTCACACCGGAGGTTTCTCTTCA 4651
Qy 67 AACATAAGGAGTTAGAAATTACAAGTAGGCAATATGCTTCTCTATATTCAGATAAAATTCATT 126
Db 4650 AACATAAGTAGTTAGAAATTACAAGTAGGCAATATGCTTCTCTATATTCAGATAAAATTCATT 4591
Qy 127 TCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGAAAAAGAAATGATAGCTATAT 186
Db 4590 TCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGAAAAAGAAATGATAGCTATAT 4532
Qy 187 TAAAGCAGATATTCATTACAATACCATGTAGACATAGCAATATTTTGGCATCATTTCT 246
Db 4531 TAAAGCAGATATTCATTACAATACCATGTAGACATAGCAATATTTTGGCATCATTTCT 4472
Qy 247 GTCCGCTCAGTAGGCGGTTCCTCTGTTAGGCGCTTTGGAGAGTACCATCTATCTAAG 306
Db 4471 GTCCGCTCAGTAGGCGGTTCCTCTGTTAGGCGCTTTGGAGAGTACCATCTATCTAAG 4412
Qy 307 ATGAGGAATCTCTGGAAGCGGGATGAGGTGCGTCTTTTACGCTGAACCCACAC 366
Db 4411 ATGAGGAATCTCTGGAAGCGGGATGAGGTGCGTCTTTTACGCTGAACCCACAC 4352
Qy 367 AGGAAATCTGAGCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATCTCTGGTCA 426
Db 4351 AGGAAATCTGAGCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATCTCTGGTCA 4292
Qy 427 ATGAAGTGAATTCCTATTTTCNGGGG 453
Db 4291 ATGAAGTGAATTCCTATTTTCNGGGG 4265
```

## RESULT 12

```
GI9707
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human STS A001W14, sequence tagged site.
105 bp
DNA
linear
STS 24-JUL-1996
GI9707.1
GI:1340278
STS: STS sequence; primer; sequence tagged site.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 105)
Adams M.D.
Unpublished (1996)
```

## AUTHORS

## COMMENT

Contact: Mark Adams  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Email: mdadams@tigr.org

Primer A: ACAATACCATGTAGAGACATAAG  
Primer B: ATCTTAGATAGTGGTACTCTCC  
STS size: 105  
PCR Profile:

Denaturation: 96C 5min  
Anneal: 54C 30sec  
Extend: 72C 30sec

```

Denature: 95C 30sec
FinalExtnd: 72C 5min
Cycles: 30

Protocol:
  GenomicDNA: 25 ng
  Primer: 0.43 uM each
  dNTPs: 230 uM each
  AmpliTaq: 0.5 units
  TagStart Ab: 0.5 units
  Total Volume: 10 ul

Buffer:
  Tris-HCl pH8.8: 100 mM
  KCl: 500 mM
  MgCl2: 20 mM
  Triton X-100: 1%
  Concentration: 10X

Prepared with primer pairs derived from THCS1137; GenBank Accession
Numbers-- F09140, F10434, T03688, T15930, T19338, T32236, T33356,
T33873.

FEATURES
    source
    1..105
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
    STS
    1..23
        primer_bind
        primer_bind
        complement (83..105)

Query Match 23.1%; Score 105; DB 11; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.6e-15;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 ACAATACCATGTAGACATAGCAATATTTGGCATCATCTTGTCCGCTCAGTAGCGCG 263
Db 1 ACAATACCATGTAGACATAGCAATATTTGGCATCATCTTGTCCGCTCAGTAGCGCG 60

QY 264 TGTTCCTCTGTGTAGGCGCTTTGGAGAGTACCATCTATCTAGAT 308
Db 61 TGTTCCTCTGTGTAGGCGCTTTGGAGAGTACCATCTATCTAGAT 105

RESULT 13
AC117965/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-366B11, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC117965
VERSION
HTG: HTGS_PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 180187)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,

```

```

Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., King,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemelehen,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Prannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savary,G., Scherer,S., Scott,C., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
Sneed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,N., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wlaczek,R., Wooden,H., Worley,K.,
Wright,D., Wright,J., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,X., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 180187)
Worley,K.C.
Direct Submission
Submitted (12-APR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 180187)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23811557.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GUOM
Center clone name: CH230-366B11
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 169631 bases at least Q40
Consensus quality: 171486 bases at least Q30

```

Consensus quality: 172615 bases at least Q20  
 Estimated insert size: 175046; sum-of-contigs estimation  
 Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

\*\*\*\*\*  
 \* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 consists of 2 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

\* 1 178750: contig of 178750 bp in length  
 \* 178751 178850: gap of unknown length  
 \* 178851 180187: contig of 1337 bp in length.

FEATURES  
 source  
 1..180187  
 /organism="Rattus norvegicus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10116"  
 /clone="CH230-366811"  
 misc\_feature  
 1..1502  
 /note="wgs\_contig"  
 misc\_feature  
 172128..173041  
 /note="clone\_boundary"  
 clone\_end:17  
 site:  
 end\_sequence:B2121917"  
 176880..178750  
 /note="wgs\_end\_extension"  
 clone\_end:17"

ORIGIN  
 Query Match 18.9%; Score 86.2; DB 2; Length 180187;  
 Best Local Similarity 71.5%; Pred. No. 1.2e-10;  
 Matches 143; Conservative 0; Mismatches 49; Indels 8; Gaps 2;  
 Qy 9 AAGTTTAAATGACACAAATTAATATATCTGTATATCTCACCGGAGNTTCTCTCAA 68  
 Db 163149 AGTTTAAATAGACACAAATTAATATATCTGTATATCTCACCGGAGNTTCTCTTAA 163090  
 Qy 69 CATAAGGAGTTAGAAATACAAAGTAGGCATATGCTTCTTATATTCAGATAAATTCATTC 128  
 Db 163089 CATAAGAG-----TTACAATTAAGTATATGCTTCTTATATTCAGATAAATTCATTC 163037  
 Qy 129 GATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATAGCTATATTA 188  
 Db 163036 TATTAATTAATTCATTTAGATTCGGAAGGGGAAATGGAAGGAACG-TCGAGATGTT 162978  
 Qy 189 AAGCAGATATTCATTACCAAT 208  
 Db 162977 AAGCAGATACACACAAAT 162958  
 RESULT 14  
 RN073458/c 5136 bp mRNA linear ROD 26-MAR-1998  
 LOCUS  
 DEFINITION Rattus norvegicus protein tyrosine phosphatase (PTPNE6) mRNA,  
 complete cds.  
 ACCSSION U73458  
 VERSION U73458.1 GI:1657945  
 KEYWORDS  
 Rattus norvegicus (Norway rat)  
 ORGANISM  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 5136)  
 REFERENCE  
 AUTHORS Fitzgerald,L.R., Walton,K.M., Dixon,J.E. and Largent,B.L.  
 TITLE PTP NE-6: a brain-enriched receptor-type protein tyrosine  
 phosphatase with a divergent catalytic domain

J. Neurochem. 68 (5), 1820-1829 (1997)  
 97263591  
 MEDLINE  
 PUBMED 9109506  
 REFERENCE  
 2 (bases 1 to 5136)  
 AUTHORS Fitzgerald,L.R., Walton,K.M., Dixon,J.E. and Largent,B.L.  
 TITLE Direct Submission  
 JOURNAL Submitted (04-OCT-1996) CNS Diseases Research, DuPont Merck  
 Research Labs, Experimental Station E400, Wilmington, DE 19880, USA  
 FEATURES  
 Location/Qualifiers  
 source  
 1..5136  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 gene  
 1..5136  
 /gene="PTPNE6"  
 CDS  
 112..3126  
 /gene="PTPNE6"  
 /function="receptor-type"  
 /codon\_start=1  
 /product="protein tyrosine phosphatase"  
 /protein\_id="AAC08036.1"  
 /db\_xref="GI:1657946"  
 /translation="MGLPLPLLLPLPPALPAPASARGQLPRLGLCFEDGL  
 CGSLETGVNDGVFGRCKVPALDTRYEVFSGALLHRIILQKLSRTGFTWQDDYTOR  
 VIAQELSNLPKAYLWHEEASPARSLQONADNEKWFSEVALAKTRRYLPYLELL  
 SOAPTANAHPRIDHETRVKGEDSPENILTYAHTSALTYPPTATKRVYDNLRLPLS  
 RLQDELSPKVDSDIDKQKLIAGAYTAQPPGENDPEPRYLVHSPMRAPRPPAAPA  
 LSORWPLPGDSKDSLSDGDDTLRLSKLQLOQAEVDRGLSKLEQADSIAQAIOS  
 DPVEGSESHGRGAEQRLQADAPPEMLQDHLPEVDDPAAYKEVRSYKSLDKGLLK  
 DHGSPLLPEAPLLEKSRAMKKEQPEEVLSSEETAGVEHVKSYSKLDLRRKPN  
 SEFPWRLEDFQFRAPEVWEDEQNLAAQSPGSGGLQLEVPSEEEQQYIITGNV  
 PLSPEKGQMLDEVAHLRVFSSFDVAVKLVGPAVIFKVSANIQMTTADLVKAAVDN  
 KBLEKATGLTIQSGIRPKGKLPLPHEEQEDSTFKFVLTFLSIACILAVLLASSL  
 AYCLRNSHYKLKELGSLGADPSADATEAVOELCRORMAVRPDQHSSEGHPTSRINSV  
 SSQSDGPMPSPARSTSWSEPAQSNWDISTGHMLAYMEDHLKNQLEKEWEA  
 LCATQABPDSLSVAQREENAPKRSALVITDHSRIILKSENSHNSDYNASIMDH  
 DPNPAYIATQGPFPATVDFWQMVSEGCACVIMLTPLSENGVRQCHHYMPDGSNV  
 YHYEVNLVSEHIWCDFLRSYLNQTNTRTVPQPHLSWYDQGVSPSSRLSD  
 FRKVNKCYGRSCPIIVHCSGDAGRSCTVLIDMLNKNMAKGAKEIDIAATLEHLRD  
 QRPQWVQTKQEFALTAVAEVNAILKALPQ"

ORIGIN  
 Query Match 18.8%; Score 85.6; DB 10; Length 5136;  
 Best Local Similarity 71.3%; Pred. No. 1.4e-10;  
 Matches 129; Conservative 0; Mismatches 45; Indels 7; Gaps 1;  
 Qy 9 AAGTTTAAATGACACAAATTAATATATTTGTATATCTCACCGGAGNTTCTCTCAA 68  
 Db 4567 AGGTTTAAATAGGACACAAATTAATATATCTGTATATCACCGGAGGTTTTCTCTTAA 4508  
 Qy 69 CATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTTATATTCAGATAAATTCATTC 128  
 Db 4507 CATAAGAG-----TTACAATTAAGTATATGCTTCTTATATTCAGATAAATTCATTC 4455  
 Qy 129 GATTAATTAATTCAGATAGAGAGTAATTTTCGAAAAGAAATGATAGCTATATTA 188  
 Db 4454 TATTAATTAATTCATTTAGATTCGCGGAAGGGGAAATGGAAGGAACGTCGAGATGTTA 4395  
 Qy 189 A 189  
 Db 4394 A 4394  
 RESULT 15  
 AC110936  
 LOCUS  
 DEFINITION Rattus norvegicus clone CH230-188N19, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 4 unordered pieces.  
 ACCESSION AC110936  
 VERSION AC110936.6 GI:25072900  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ENRICHED.  
 SOURCE Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 222251)  
Muzny, D., Marie, Metzker, M., Lee, A., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Fallis, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowig, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulseghe, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L., Nankervils, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokeme, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, L., Pu, M., Quirroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Speed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valae, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G., and Gibbs, R. A.

TITLE  
JOURNAL

Direct Submission

REFERENCE  
AUTHORS

2 (bases 1 to 222251)  
Worley, K.C.

TITLE  
JOURNAL

Submitted (17-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 222251)  
Rat Genome Sequencing Consortium.

REFERENCE  
AUTHORS

Direct Submission

TITLE  
JOURNAL

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Nov 19, 2002 this sequence version replaced gi:23270153.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas

assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GOFU

Center clone name: CH230-188N19

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 193700 bases at least Q40

Consensus quality: 196702 bases at least Q30

Consensus quality: 198783 bases at least Q20

Estimated insert size: 196105; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved

\* 1 213437: contig of 213437 bp in length

\* 213438 213537: gap of unknown length

\* 213538 218989: contig of 5452 bp in length

\* 218990 219089: gap of unknown length

\* 219090 220660: contig of 1571 bp in length

\* 220661 220760: gap of unknown length

\* 220761 222251: contig of 1491 bp in length.

## FEATURES

## source

1..222251

/organism="Rattus norvegicus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10116"

/clone="CH230-188N19"

1..1121

/note="wgs\_contig"

16434..18286

/note="wgs\_contig"

27659..28735

/note="wgs\_contig"

128775..130020

/note="wgs\_contig"

174053..175360

/note="wgs\_contig"

183232..184558

/note="wgs\_contig"

213538..215409

/note="wgs\_contig"

## ORIGIN

Query Match 16.9%; Score 76.8; DB 2; Length 222251;

Best Local Similarity 80.6%; Pred. No. 2.7e-08;

Matches 104; Conservativity 0; Mismatches 18; Indels 7; Gaps 1;

QY 9 AAGTTTAATATGACACAAATTAATATATTTGATATCTCACACCGAGNTTCTTTCAA 68

Db 127895 AGTTTAATAGACACAAATTAATATCTGATATATCACACCGAGGTTTTCCTTTAA 127954

QY 69 CATAGAGGATTAGAATTAAGTAGGCGATATGCTTCTTATATTCAGATAAATTCATTC 128



Db 127955 CATAAGAAG-----TTACAAATTAAGTATATGCTTCCTATATTCAGATAAATTCATTTTC 128007

Qy 129 GATTAATTA 137

Db 128008 TATTAATTA 128016  
|||||

Search completed: May 5, 2005, 03:58:33  
Job time : 2231.09 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 20:02:41 ; Search time 314.957 Seconds  
(without alignment)  
8551.893 Million cell updates/sec

Title: US-08-731-499-8  
Perfect score: 455  
Sequence: 1 GAATCAGAGTTTAATG.....ATTGCTCTATTTCNGGGGT 455

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_16Dec04.\*  
1: Geneseqn1980s.\*  
2: Geneseqn1990s.\*  
3: Geneseqn2000s.\*  
4: Geneseqn2001as.\*  
5: Geneseqn2001bs.\*  
6: Geneseqn2002as.\*  
7: Geneseqn2002bs.\*  
8: Geneseqn2003as.\*  
9: Geneseqn2003bs.\*  
10: Geneseqn2003cs.\*  
11: Geneseqn2003ds.\*  
12: Geneseqn2004as.\*  
13: Geneseqn2004bs.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	99.6	455	AAV09022	Homo sapi
C 2	439	96.5	1258	AAH34753	Human col
C 3	426.4	93.7	2328	AAT95227	Human isl
C 4	426	93.6	4699	ADD14680	Human SRC
C 5	426	93.6	4767	ADN95437	Human BEC
C 6	414.6	91.1	4719	ADP18660	Human TAT
C 7	254.8	56.0	282	ABV95893	Human pan
C 8	225.2	49.5	4856	AAH87633	DNA encod
9	120.6	26.5	121	ADF86738	Single nu
10	114.6	25.2	121	ADF86737	Single nu
11	109.6	24.1	121	ADF86735	Single nu
12	107.6	23.6	121	ADF86734	Single nu
13	92.6	20.4	121	ADF86736	Single nu
C 14	64.4	14.2	455	AAH35240	Human col
C 15	54.6	12.0	538	ADQ53174	Novel can
16	49.4	10.9	51	AAH29285	Human SNP
C 17	45.4	10.0	2000	ADA71938	Rice gene
18	44.2	9.7	6811	ABH34540	Human net
19	44.2	9.7	6811	ABH70263	Chemical
20	44.2	9.7	6811	ADS99801	Bisulphit

C 21	44.2	9.7	11422	6	ABK39937	Human che
C 22	44.2	9.7	11422	6	ABL32219	Human imm
C 23	44.2	9.7	37515	6	ABQ66997	Human ang
24	44	9.7	2000	8	ADA71938	Rice gene
25	43	9.5	1246	2	AZ42154	Human nor
26	43	9.5	2117	2	AAT61258	Human cDN
27	43	9.5	2815	3	AAT21830	Human bre
28	43	9.5	3323	13	ADR67210	Human bla
29	43	9.5	3323	13	ADR66257	Human pro
30	43	9.5	3323	13	ADR66599	Human pro
C 31	43	9.5	61020	4	AA546787	Human su
32	42.6	9.4	51	4	AAI29288	Human SNP
33	42.6	9.4	344	4	AAH93307	Human SNP
34	42	9.2	51	4	AAI29282	Human SNP
35	42	9.2	1132	3	AAF12929	Human SNP
C 36	42	9.2	8056	8	ABZ10246	Haematopo
C 37	42	9.2	13574	6	ABL33317	Human imm
C 38	41.6	9.1	2849	4	ABL14726	Drosophil
C 39	41.4	9.1	110000	13	ABD32966	Continuatio
C 40	41.2	9.1	6988	6	ABL34441	Human imm
41	41	9.0	11922	3	AAA70187	Plasmodiu
C 42	40.6	8.9	794	6	ABQ42952	Oligonuc
43	40.6	8.9	794	6	ABQ42953	Oligonuc
C 44	40.4	8.9	337	12	ADO55043	Gene #140
45	40.4	8.9	700	4	AAH93026	Human inf

ALIGNMENTS

RESULT 1  
AAV09022  
ID AAV09022 standard; cDNA; 455 BP.

AC AAV09022;  
DT 21-JUL-1998 (first entry)  
DE Homo sapiens 20ql3 amplicon 20sa7 transcript.  
KW 20ql3 amplicon; chromosome 20; tumour; detection;  
KW chromosomal abnormalities; probe; gene therapy; antisense inhibition;  
KW treatment; age-related macular degeneration; retinitis pigmentation;  
KW Leber's congenital amaurosis; BEM-1; ds.

OS Homo sapiens.  
PN WO9802539-A1.  
XX 22-JAN-1998.  
XX 15-JUL-1997; 97WO-US012343..

PR 15-JUL-1996; 96US-00680395.  
PR 16-OCT-1996; 96US-00731499.  
PR 17-JAN-1997; 97US-00785532.  
PA (REGC ) UNIV CALIFORNIA.

PI Gray JW, Collins CC, Hwang S, Godfrey T, Kowbel D, Rommens J;  
XX WPI; 1998-110587/10.

XX New sequences from the 20ql3 amplicon - used for detecting chromosomal  
PT abnormalities, particularly tumours, and for developing products for  
PT treating diseases.

PS Claim 1; Page 63-64; 91pp; English.

XX The sequence is that of a cDNA sequence 20sa7 which was isolated from the  
CC 20ql3 amplicon. It is a homologue of the rat gene BEM-1 and can be used  
CC as a probe for the detection of chromosomal abnormalities at 20ql3. It  
CC and other sequences isolated from the 20ql3 amplicon are consistently

CC amplified in primary tumours. These sequences are useful as probes or as  
CC probe targets for monitoring the relative copy number of corresponding  
CC sequences from a biological sample such as tumour cells. The sequences  
CC can also be used in therapeutic applications for modulating the  
CC expression of the endogenous gene or the activity of the gene product.  
CC Examples of therapeutic approaches include antisense inhibition of gene  
CC expression, gene therapy, and monoclonal antibodies that specifically  
CC bind the gene products. The products can also be used in the treatment of  
CC other diseases, e.g. age-related macular degeneration, Leber's congenital  
CC amaurosis and retinitis pigmentation  
XX  
SQ Sequence 455 BP; 138 A; 87 C; 98 G; 130 T; 0 U; 2 Other;

Query Match 99.6%; Score 453; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 5e-118;  
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GAAATCAGAGTTTAATATGACACAAATTAATATATTTGTATATCTTCACACCGAGNTTC 60  
DB 1 GAAATCAGAGTTTAATATGACACAAATTAATATATTTGTATATCTTCACACCGAGNTTC 60  
QY 61 TCTTCAACATAGAGTTAGAAATTTACAGTAGGATAGGATATGCTTCTATATTCAGATAAA 120  
DB 61 TCTTCAACATAGAGTTAGAAATTTACAGTAGGATAGGATATGCTTCTATATTCAGATAAA 120  
QY 121 TTCAATTTCCGATTAATTAATTTCCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 180  
DB 121 TTCAATTTCCGATTAATTAATTTCCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 180  
QY 181 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATPAAGCAATATTTTGGCAT 240  
DB 181 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATPAAGCAATATTTTGGCAT 240  
QY 241 CATTCGTCCGCTCAGTAGGCCGTGTCCCTCTGGTAGGGCCTTTGGAGATACCATCTA 300  
DB 241 CATTCGTCCGCTCAGTAGGCCGTGTCCCTCTGGTAGGGCCTTTGGAGATACCATCTA 300  
QY 301 TCTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGTGGTGTCTTCTACGCTGAACC 360  
DB 301 TCTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGTGGTGTCTTCTACGCTGAACC 360  
QY 361 CCACACAGGAAATCTCAGCCCAACACAGCTGCTCTCGCGCGCCTTCCATGTGATCATCC 420  
DB 361 CCACACAGGAAATCTCAGCCCAACACAGCTGCTCTCGCGCGCCTTCCATGTGATCATCC 420  
QY 421 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGGT 455  
DB 421 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGGT 455

RESULT 2  
AAH34753/c  
ID AAH34753 standard; cDNA; 1258 BP.  
XX  
AC AAH34753;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1835.  
XX  
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; chromosome 7; 88.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US026524.  
XX  
PR 29-SEP-1999; 99US-0157137P.  
PR 03-NOV-1999; 99US-0163280P.

(HUMA-) HUMAN GENOME SCI INC.  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX WPI; 2001-235357/24.  
DR P-PSDB; AAG75348.  
XX  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers.  
XX  
PS Claim 1; Page 3372-3373; 9803pp; English.  
XX  
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the  
CC proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene therapy  
CC and vaccine production. N and P may be used in the prevention, diagnosis  
CC and treatment of diseases associated with inappropriate P expression. For  
CC example, N and P may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of P by expressing inactive proteins or to  
CC supplement the patients own production of P. Additionally, N may be used  
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids  
CC into a host cell and culturing the cell to express the proteins. N and P  
CC can be used in the prevention, diagnosis and treatment of colorectal  
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent  
CC sequences used in the exemplification of the present invention. N.B.  
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at  
CC time of publication, meaning no sequences are present for SEQ ID NO:1027  
CC to 1052, 7921 and 7922  
XX  
SQ Sequence 1258 BP; 345 A; 325 C; 302 G; 284 T; 0 U; 2 Other;  
Query Match 96.5%; Score 439; DB 4; Length 1258;  
Best Local Similarity 99.3%; Pred. No. 6.9e-114;  
Matches 450; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 1 GAAATCAGAGTTTAATATGACACAAATTAATATATTTGTATATCTTCACACCGAGNTTC 60  
DB 1213 GAAATCAGAGTTTAATATGACACAAATTAATATATTTGTATATCTTCACACCGAGNTTC 1154  
QY 61 TCTTCAACATAGAGTTAGAAATTTACAGTAGGATATGCTTCTATATTCAGATAAA 120  
DB 1153 TCTTCAACATAGAGTTAGAAATTTACAGTAGGATATGCTTCTATATTCAGATAAA 1094  
QY 121 TTCAATTTCCGATTAATTAATTTCCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 180  
DB 1093 TTCAATTTCCGATTAATTAATTTCCAGATAGAGAGAAGTAATTTTCGGAAGAAGAAATCATAG 1035  
QY 181 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATPAAGCAATATTTTCGSCAT 240  
DB 1034 CTATATTAAGCAGATATTCATTTACAAATACCATGTAGAGACATPAAGCAATATTTTCGSCAT 975  
QY 241 CATTCGTCCGCTCAGTAGGCCGTGTCCCTCTGGTAGGGCCTTTGGAGATACCATCTA 300  
DB 974 CATTCGTCCGCTCAGTAGGCCGTGTCCCTCTGGTAGGGCCTTTGGAGATACCATCTA 915  
QY 301 TCTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGTGGTGTCTTCTACGCTGAACC 360  
DB 914 TCTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGTGGTGTCTTCTACGCTGAACC 855  
QY 361 CCACACAGGAAATCTCAGCCCAACACAGCTGCTCTCGCGCGCCTTCCATGTGATCATCC 420  
DB 854 CCACACAGGAAATCTCAGCCCAACACAGCTGCTCTCGCGCGCCTTCCATGTGATCATCC 795  
QY 421 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGG 453  
DB 794 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGG 762  
RESULT 3  
AAT95227/c

ID AAT95227 standard; cDNA; 2328 BP.  
XX AAT95227;  
AC  
XX 27-MAR-1998 (first entry)  
DT  
XX Human islet cell antigen 1851 partial cDNA clone WK121315.  
DE  
XX Islet cell antigen 1851; autoantigen; protein tyrosine phosphatase;  
KW insulin-dependent diabetes mellitus; IDDM; diagnosis; therapy; human; ds.  
XX  
XX Homo sapiens.  
OS  
XX WO9732984-A1.  
FN  
XX 12-SEP-1997.  
PD  
XX 05-MAR-1997; 97WO-US003532.  
PF  
XX 06-MAR-1996; 96US-0012927P.  
PR  
XX 15-OCT-1996; 96US-0027540P.  
PR  
XX (ZYMO) ZYMOGENETICS INC.  
PA (UNIW) UNIV WASHINGTON.  
PA  
XX Kindsvogel W, Jelinek L, Sheppard PO, Hagopian W, Lagasse J;  
PI WPI; 1997-457535/42.  
DR  
XX Mammalian pancreatic islet cell antigen and related DNA - used to detect  
PT auto-antibodies indicative of insulin-dependent diabetes or pre-  
PT disposition to it.  
PT  
XX Example 5; Page 114-115; 134pp; English.  
PS  
XX This cDNA sequence comprises clone WK121315 a partial cDNA clone of novel  
CC human islet cell antigen 1851. It was isolated from human insulinoma cDNA  
CC by PCR amplification (see AAT95224-25). Other partial clones (see  
CC AAT95226 and AAT95228) were similarly obtained. Subsequent 5' RACE  
CC yielded a claimed sequence (see AAT95221) for islet cell antigen 1851  
CC (see AAW35297). This antigen forms an immune complex with an autoantibody  
CC found in patients at risk of, or predisposed to, insulin-dependent  
CC diabetes mellitus (IDDM), and can be used in diagnostic and therapeutic  
CC methods  
XX  
SQ Sequence 2328 BP; 597 A; 599 C; 595 G; 537 T; 0 U; 0 Other;  
  
Query Match 93.7%; Score 426.4; DB 2; Length 2328;  
Best Local Similarity 98.9%; Pred. No. 3.3e-110;  
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
  
QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNNTT 59  
DB 2282 GAATCAGAAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACGGGTTT 2223  
  
QY 60 CTCCTTCAACATAAGGAGTGAATAATCAAGTAGGCATATGCTTCTTATATTCAGATAA 119  
DB 2222 CTCCTTCAACATAAGGAGTGAATAATCAAGTAGGCATATGCTTCTTATATTCAGATAA 2163  
  
QY 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAATGATA 179  
DB 2162 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAATGATA 2104  
  
QY 180 GCTATATTAAAGCAGATATTCTAATCAATACCATCTAGACATAGCAATATTTTGGCA 239  
DB 2103 GCTATATTAAAGCAGATATTCTAATCAATACCATCTAGACATAGCAATATTTTGGCA 2044  
  
QY 240 TCATTTCTGCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299  
DB 2043 TCATTTCTGCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 1984  
  
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGAGGTGCGTTTTTCTACGCTGAAC 359  
|||||

Db 1393 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGAGGTGCGTTTTTCTACGCTGAAC 1924  
Qy 360 CCCACACAGAAATCTGCAGCCCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATC 419  
|||||  
Db 1923 CCCACACAGAAATCTGCAGCCCCACACAGCTGCTCTGCGCGCTTCCATGTGATCATC 1864  
Qy 420 CTGCTCAATGAAGTAATTCCTTATTCNGGGG 453  
|||||  
Db 1863 CTGCTCAATGAAGTAATTCCTTATTCNGGGG 1830  
|||||  
RESULT 4  
ADD14680/C  
ID ADD14680 standard; cDNA; 4699 BP.  
XX  
XX AC  
XX ADD14680;  
XX  
XX 01-JAN-2004 (first entry)  
XX  
XX Human src biomarker polynucleotide SEQ ID NO:74.  
XX  
XX predictor set; protein tyrosine kinase activity modulator;  
XX protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;  
XX gene therapy; drug sensitivity; genetic profile; cancer; human; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO2003062395-A2.  
XX  
XX 31-JUL-2003.  
XX  
XX 17-JAN-2003; 2003WO-US001981.  
XX  
XX 18-JAN-2002; 2002US-0350061P.  
XX  
XX (BRIM) BRISTOL-MYERS SQUIBB CO.  
XX  
XX Huang F, Fairchild CR, Lee FY, Shaw P;  
XX WPI; 2003-636735/60.  
XX P-PSDB; ADD14084.  
XX  
XX New polynucleotides and polypeptides for predicting the activity of  
PT compounds that interact with protein tyrosine kinases and/or protein  
PT tyrosine kinase pathways.  
XX  
XX Claim 2; SEQ ID NO 74; 139pp; English.  
PS  
XX The present invention describes a predictor set comprising a plurality of  
XX polynucleotides or polypeptides whose expression pattern is predictive of  
XX the response of cells to treatment with a compound that modulates protein  
XX tyrosine kinase activity or members of the protein tyrosine kinase  
XX pathway. Also described: (1) predicting whether a compound is capable of  
XX modulating the activity of cells, comprising obtaining a sample of cells,  
XX determining whether the cells express a plurality of markers, and  
XX correlating the expression of the markers to the compound's ability to  
XX modulate the activity of the cells; (2) a plurality of cell lines for  
XX identifying polynucleotides and polypeptides whose expression levels  
XX correlate with compound sensitivity or resistance of cells associated  
XX with a disease state; and (3) identifying polynucleotides and  
XX polypeptides that predict compound sensitivity or resistance of cells  
XX associated with a disease state, comprising subjecting the plurality of  
XX cell lines to one or more compounds, analysing the expression pattern of  
XX a microarray of polynucleotides or polypeptides, and selecting  
XX polynucleotides or polypeptides that predict the sensitivity or  
XX resistance of cells associated with a disease state by using the  
XX expression pattern of the microarray. The polynucleotides and  
XX polypeptides have cytostatic activities, and can be used in gene therapy.  
XX The polynucleotides and polypeptides are useful in predicting the  
XX activity of compounds that interact with protein tyrosine kinases and/or  
XX protein tyrosine kinase pathways. These may be used in determining drug  
XX sensitivity in patients to allow the development of individualized  
XX genetic profiles which aid in treating diseases and disorders (e.g.



```
RESULT 6
ADP18660/c
ID ADP18660 standard; cDNA; 4719 BP.
XX
AC ADP18660;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human TAT418 cDNA used to treat cancer SeqID 20.
XX
KW tumour; growth inhibitory; cytotoxic; cytostatic; gene therapy; cancer;
KW human; TAT; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2004045516-A2.
XX
PD 03-JUN-2004.
XX
PF 13-NOV-2003; 2003WO-US036298.
XX
PR 15-NOV-2002; 2002US-0426847P.
PR 06-DEC-2002; 2002US-0431250P.
PR 31-DEC-2002; 2002US-0437344P.
XX
PA (GETH ) GENENTECH INC.
XX
PI Ashkenazi AJ, Frantz G, Goddard A, Gonzalez L, Gurney AL;
PI Polakis P, Polson A, Wood WI, Wu TD, Zhang Z;
XX
DR WPI; 2004-420515/39.
DR P-PSDB; ADP18661.
XX
PT New antibody that binds to tumor-associated antigenic target polypeptide
PT (TAT), useful in preparing a composition for diagnosing or treating
PT tumor.
XX
PS Claim 1; SEQ ID NO 20; 183pp; English.
XX
CC This invention relates to novel isolated antibodies that are useful for
CC the diagnosis and treatment of a tumour. Specifically, it refers to DNA
CC that is detectably labelled and conjugated to a growth inhibitory or
CC cytotoxic agent and comprises toxins, antibiotics, radioactive isotopes
CC or nucleolytic enzymes. The present invention refers to the toxin
CC maytansinoid or calicheamicin, where the host cell is one of chinese
CC hamster ovary (CHO), yeast or Escherichia coli cell. Furthermore, these
CC pharmaceutical compositions are cytostatic and can be used for gene
CC therapy purposes to treat various cancers. This polynucleotide sequence
CC is a human TAT cDNA sequence of the invention.
XX
SQ Sequence 4719 BP; 1057 A; 1404 C; 1340 G; 918 T; 0 U; 0 Other;

Query Match 91.1%; Score 414.6; DB 12; Length 4719;
Best Local Similarity 98.2%; Pred. No. 9.6e-107;
Matches 439; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

QY 8 GAAAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNITCTCTTCA 66
DB |||||||
DB 4710 GAAGTTTAAATGACACAAATTAATATATTTGTATATCTCACACCGAGGTTCTTCA 4651
QY 67 AACATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAATTCATT 126
DB |||||||
DB 4650 AACATAAGTAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAATTCATT 4591
QY 127 TCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGTATAT 186
DB |||||||
DB 4590 TCGATTAAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGTATAT 4532
QY 187 TAAACGACATATTCATTACATACCATCATGATAGACATAGCAATATTTTGGCATCATCT 246
DB |||||||
DB 4531 TAAACGACATATTCATTACATACCATCATGATAGACATAGCAATATTTTGGCATCATCT 4472
```

## RESULT 7

ABV95893/c

ID ABV95893 standard; cDNA; 282 BP.

XX AC ABV95893;

XX XX

XX 14-JAN-2003 (first entry)

XX XX

DE Human pancreatic cancer expressed cDNA SEQ ID NO 1301.

XX XX

KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;

XX KW

XX Cytostatic; tumour; gene; ss.

XX XX

XX Homo sapiens.

XX OS

XX WO200260317-A2.

XX PN

XX PD

XX 08-AUG-2002.

XX XX

XX 30-JAN-2002; 2002WO-US002781.

XX XX

XX 30-JAN-2001; 2001US-0265305P.

XX PR

XX 31-JAN-2001; 2001US-0265682P.

XX PR

XX 09-FEB-2001; 2001US-0267568P.

XX PR

XX 21-MAR-2001; 2001US-0278651P.

XX PR

XX 28-APR-2001; 2001US-0287112P.

XX PR

XX 16-MAY-2001; 2001US-0291631P.

XX PR

XX 12-JUL-2001; 2001US-0305484P.

XX PR

XX 20-AUG-2001; 2001US-0313999P.

XX PR

XX 27-NOV-2001; 2001US-0333626P.

XX XX

XX (CORI-) CORIXA CORP.

XX PA

XX XX

XX Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;

XX PI

XX WPI; 2002-627435/67.

XX DR

XX XX

XX New isolated polynucleotide and pancreatic tumor polypeptides, useful for

XX PT

XX diagnosing, preventing and/or treating cancer, particularly pancreatic

XX PT

XX cancer.

XX XX

XX Claim 1; SEQ ID NO 1301; 300pp + Sequence Listing; English.

XX PS

XX CC

XX

The invention relates to an isolated polynucleotide (I) comprising: (a) any of a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b) complements of (a); (c) sequences consisting of at least 20 contiguous residues of (a); (d) sequences that hybridize to (a), under moderately stringent conditions; (e) sequences having at least 75% or 90% identity to (a); or (f) degenerate variants of (a). Polypeptides (ABP68536-ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer in a patient and compositions comprising polypeptides, polynucleotides, antibodies, fusion proteins, T cell populations and antigen presenting cells expressing the polypeptide are useful in treating pancreatic cancer and stimulating an immune response. The polynucleotides can be used as probes or primers for nucleic acid hybridisation, in the design and preparation of ribozyme molecules for inhibiting expression of the tumour

CC polypeptides and proteins in the tumour cells, in vaccines and for gene  
CC therapy. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 282 BP; 92 A; 49 C; 44 G; 97 T; 0 U; 0 Other;  
Query Match 56.0%; Score 254.8; DB 6; Length 282;  
Best Local Similarity 98.2%; Pred. No. 5.8e-62;  
Matches 278; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
QY 11 GTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNNTCTCTCAAAAC 69  
Db |||||||  
QY 282 GTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGAGTTCTCTCAAAAC 223  
Db |||||||  
QY 70 ATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCG 129  
Db |||||||  
QY 222 ATAAGTAGTGAAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCG 163  
Db |||||||  
QY 130 ATTAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGCTATATTTAA 189  
Db |||||||  
QY 162 ATTAATTAATTCAGATAGAGAGAGTAATTTTCGAAAAGAAATGATAGCTATATTTAA 104  
QY 190 AGCAGATATTCATTAACAATACCATGATAGAGACATAGCAATATTTTGGCATCATTTCTGTC 249  
Db |||||||  
QY 103 AGCAGATATTCATTAACAATACCATGATAGAGACATAGCAATATTTTGGCATCATTTCTGTC 44  
QY 250 CGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGT 292  
Db |||||||  
QY 43 CGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGT 1

RESULT 8  
AAS87633/c  
ID AAS87633 standard; cDNA; 4856 BP.  
XX  
AC AAS87633;  
DT 13-FEB-2002 (first entry)  
DE DNA encoding novel human diagnostic protein #23437.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG23446.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 23437; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (III) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
CC coding sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 4856 BP; 1082 A; 1454 C; 1374 G; 946 T; 0 U; 0 Other;  
Query Match 49.5%; Score 225.2; DB 5; Length 4856;  
Best Local Similarity 82.4%; Pred. No. 4.4e-53;  
Matches 394; Conservative 0; Mismatches 60; Indels 24; Gaps 11;  
QY 1 GAAATCAGAAGTTTAAATATATGACACAAATTAATATTTGTATATCTCACACCGG-AGNNTT 59  
Db |||||||  
QY 4833 GAAATCAGAAGTTTAAATATATGACACAAATTAATATTTGTATATCTCACACCGGAAGTTT 4774  
Db |||||||  
QY 60 CTCCTTCAACATACAGAGTTA-GAAATTCAGATAGGATAGCTTCTCTATATTCAGATA 118  
Db |||||||  
QY 4773 CTCCTTCAACATACAGAGTTTAAATTAACAATTAACAATAGGCATATGCTTCTCTATATTCAGATA 4714  
Db |||||||  
QY 119 AATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAAAG-AAATGA 177  
Db |||||||  
QY 4713 AATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAAAGCAAATGA 4654  
Db |||||||  
QY 178 TAGCTATATTAAG--CAGATATTCATTAACAATACCATGTAGAGACATAAG--CAATATT 233  
Db |||||||  
QY 4653 TAGCTATATTAAGGCAGGATATTCATTAACAATACCATGTAGAGACATAACGGCAATATT 4594  
Db |||||||  
QY 234 TTGGGATCATTTCTGTCGGCT----CAGTAGGCGGTGTTCCCTCTGTTAGTGGGCGC---TTTGG 286  
Db |||||||  
QY 4593 TGGGGATCATTTCTGTCGGCTTTCAGTAGGCGCGTGTTCCTCTGTTAGTGGGCGCCTTTTGG 4534  
Db |||||||  
QY 287 GAGAGTACCATCTATCTAAGATGGAGGATGCTGTGGGAAGGCGGGATGGAGGTG---C 343  
Db |||||||  
QY 4533 GAGAGTACCATCTATCTAAGATGGAGGATGCTGTGGGAAGGCGGGATGGAGGTGCGC 4474  
Db |||||||  
QY 344 GTTTTCTACGCTGAACCCACACACA--GGAATCTGCAGCCCAACAGC-TGCTCTGCGC 400  
Db |||||||  
QY 4473 GTTTTCTACGCTGAACCCACACAGGAATCTGCGAGCCCAACAGCTTGCTCTGCGG 4414  
Db |||||||  
QY 401 CGCCTTCCATGTGATCATCTCT---GGTCAATGAAGTGAATTCCTATTCNGGGG 454  
Db |||||||  
QY 4413 CGCCTTTCATGTGATCATCTCTGGGTCAATGAATGAATTCCTAATTTCTGCGG 4356  
Db |||||||  
RESULT 9  
ADF86738  
ID ADF86738 standard; DNA; 121 BP.  
XX  
AC ADF86738;  
DT 26-FEB-2004 (first entry)  
DE Single nucleotide polymorphism detection human gene, SEQ ID No 321.  
XX  
XX human; single nucleotide polymorphism; microarray; side effect; gene; ds.  
XX  
XX Homo sapiens.  
OS  
XX JP2003235571-A.



XX PD 26-AUG-2003.  
 XX PF 12-FEB-2002; 2002JP-00034717.  
 XX PR 12-FEB-2002; 2002JP-00034717.  
 XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 XX DR WPI; 2003-820454/77.  
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.  
 XX PS Claim 1; SEQ ID NO 321; 704pp; Japanese.  
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given in specification, or a sequence having a base substitution. The invention further relates to: an oligonucleotide containing single nucleotide polymorphisms; a PCR primer set chosen from the combination of two DNA fragments from any one of 1220 fully defined sequences as given in specification; a labelling probe containing the SNP containing oligo; and a microarray equipped with the SNP containing oligo. The isolated human gene of the invention is useful for detecting the single nucleotide polymorphisms in human gene. The isolated human gene is also useful for diagnosis of disease and determination of side effect to a medical agent. The isolated human gene is also effective in detecting single nucleotide polymorphisms in a human gene. This polynucleotide sequence represents one of the 935 isolated polynucleotides from a human gene of the invention.

XX SQ Sequence 121 BP; 27 A; 33 C; 34 G; 26 T; 0 U; 1 Other;  
 Query Match 26.5%; Score 120.6; DB 10; Length 121;  
 Best Local Similarity 99.2%; Pred. No. 4.4e-24;  
 Matches 120; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 292 TACCATCTATCTAAGATGGAGGAATGCTGTGGAGGCGGGATGGAGTGGTGGTTTCTA 351  
 Db 1 TACCATCTATCTAAGATGGAGGAATGCTGTGGAGGCGGGATGGAGTGGTGGTTTCTA 60

Qy 352 CGTGAACCCACACAGAGAAATCTCAGCCACACAGCTGCTGTGGCGGCTTCCATG 411  
 Db 61 YGCTGAACCCACACAGAGAAATCTCAGCCACACAGCTGCTGTGGCGGCTTCCATG 120

Qy 412 T 412  
 Db 121 T 121

RESULT 10  
 ADF86737  
 ID ADF86737 standard; DNA; 121 BP.  
 AC ADF86737;  
 DT 26-FEB-2004 (first entry)  
 DE Single nucleotide polymorphism detection human gene, SEQ ID No 320.  
 KW human; single nucleotide polymorphism; microarray; side effect; gene; ds.  
 OS Homo sapiens.  
 XX JP2003235571-A.  
 XX PD 26-AUG-2003.  
 XX PF 12-FEB-2002; 2002JP-00034717.  
 XX PR 12-FEB-2002; 2002JP-00034717.

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 XX WPI; 2003-820454/77.  
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.  
 XX PS Claim 1; SEQ ID NO 320; 704pp; Japanese.  
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given in specification, or a sequence having a base substitution. The invention further relates to: an oligonucleotide containing single nucleotide polymorphisms; a PCR primer set chosen from the combination of two DNA fragments from any one of 1220 fully defined sequences as given in specification; a labelling probe containing the SNP containing oligo; and a microarray equipped with the SNP containing oligo. The isolated human gene of the invention is useful for detecting the single nucleotide polymorphisms in human gene. The isolated human gene is also useful for diagnosis of disease and determination of side effect to a medical agent. The isolated human gene is also effective in detecting single nucleotide polymorphisms in a human gene. This polynucleotide sequence represents one of the 935 isolated polynucleotides from a human gene of the invention.

XX SQ Sequence 121 BP; 25 A; 34 C; 30 G; 31 T; 0 U; 1 Other;  
 Query Match 25.2%; Score 114.6; DB 10; Length 121;  
 Best Local Similarity 98.3%; Pred. No. 2.2e-22;  
 Matches 114; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 338 AGTGGCGTTTCTAGCTGAACCCACACAGAGAAATCTCAGCCACACAGCTGCTCTG 397  
 Db 1 AGTGGCGTTTCTAGCTGAACCCACACAGAGAAATCTCAGCCACACAGCTGCTCTG 60

Qy 398 CGCGCCCTTCCATGATCATCTGCTCAATGAATGTAATGTCCTATTTCNGGGG 453  
 Db 61 YGCGCCCTTCCATGATCATCTGCTCAATGAATGTAATGTCCTATTTCNGGGG 116

RESULT 11  
 ADF86735  
 ID ADF86735 standard; DNA; 121 BP.  
 AC ADF86735;  
 DT 26-FEB-2004 (first entry)  
 DE Single nucleotide polymorphism detection human gene, SEQ ID No 318.  
 KW human; single nucleotide polymorphism; microarray; side effect; gene; ds.  
 OS Homo sapiens.  
 XX JP2003235571-A.  
 XX PD 26-AUG-2003.  
 XX PF 12-FEB-2002; 2002JP-00034717.  
 XX PR 12-FEB-2002; 2002JP-00034717.  
 XX PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
 XX WPI; 2003-820454/77.  
 XX PT Novel polynucleotide useful for detecting single nucleotide polymorphisms in human gene.  
 XX PS Claim 1; SEQ ID NO 318; 704pp; Japanese.  
 XX CC The invention relates to a novel polynucleotide isolated and purified from a human gene having any one of 935 fully defined sequences as given

CC in specification, or a sequence having a base substitution. The invention  
CC further relates to: an oligonucleotide containing single nucleotide  
CC polymorphisms; a PCR primer set chosen from the combination of two DNA  
CC fragments from any one of 1220 fully defined sequences as given in  
CC specification; a labelling probe containing the SNP containing oligo; and  
CC a microarray equipped with the SNP containing oligo. The isolated human  
CC gene of the invention is useful for detecting the single nucleotide  
CC polymorphisms in human gene. The isolated human gene is also useful for  
CC diagnosis of disease and determination of side effect to a medical agent.  
CC The isolated human gene is also effective in detecting single nucleotide  
CC polymorphisms in a human gene. This polynucleotide sequence represents  
CC one of the 935 isolated polynucleotides from a human gene of the  
CC invention.  
XX  
SQ Sequence 121 BP; 50 A; 14 C; 17 G; 39 T; 0 U; 1 Other;

Query Match 24.1%; Score 109.6; DB 10; Length 121;  
Best Local Similarity 98.4%; Pred. No. 5.7e-21;  
Matches 120; Conservative 1; Mismatches 0; Indels 1; Gaps 1;  
QY 89 AAGTAGGCATATGCTTCCTATATTCAGATAAAATTCATTTCGATTAAATTCAGATA 148  
DB 1 AAGTAGGCATATGCTTCCTATATTCAGATAAAATTCATTTCGATTAAATTCAGATA 60  
QY 149 GAGAGAAGTAATTTTCGAAAAAGAAATGATAGCTATATTAAGCAGATATTCATTACAAT 208  
DB 61 RAGAGAAGTAATTTT-GGAAAGAAATGATAGCTATATTAAGCAGATATTCATTACAAT 119  
QY 209 AC 210  
DB 120 AC 121

RESULT 12  
ADP86734  
ID ADF86734 standard; DNA; 121 BP.  
XX AC ADF86734;  
XX 26-FEB-2004 (first entry)  
XX Single nucleotide polymorphism detection human gene, SEQ ID No 317.  
DE human; single nucleotide polymorphism; microarray; side effect; gene; ds.  
XX Homo sapiens.  
XX JP2003235571-A.  
XX 26-AUG-2003.  
XX 12-FEB-2002; 2002JP-00034717.  
XX 12-FEB-2002; 2002JP-00034717.  
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
XX WPI; 2003-820454/77.  
XX Novel polynucleotide useful for detecting single nucleotide polymorphisms  
XX in human gene.

PS Claim 1; SEQ ID NO 317; 704pp; Japanese.  
XX The invention relates to a novel polynucleotide isolated and purified  
CC from a human gene having any one of 935 fully defined sequences as given  
CC in specification, or a sequence having a base substitution. The invention  
CC further relates to: an oligonucleotide containing single nucleotide  
CC polymorphisms; a PCR primer set chosen from the combination of two DNA  
CC fragments from any one of 1220 fully defined sequences as given in  
CC specification; a labelling probe containing the SNP containing oligo; and  
CC a microarray equipped with the SNP containing oligo. The isolated human  
CC gene of the invention is useful for detecting the single nucleotide

CC polymorphisms in human gene. The isolated human gene is also useful for  
CC diagnosis of disease and determination of side effect to a medical agent.  
CC The isolated human gene is also effective in detecting single nucleotide  
CC polymorphisms in a human gene. This polynucleotide sequence represents  
CC one of the 935 isolated polynucleotides from a human gene of the  
CC invention.  
XX

SQ Sequence 121 BP; 44 A; 19 C; 15 G; 42 T; 0 U; 1 Other;  
Query Match 23.6%; Score 107.6; DB 10; Length 121;  
Best Local Similarity 97.5%; Pred. No. 2.1e-20;  
Matches 118; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

QY 16 ATATGACACAATTAATATATTTGTATATCTCACACCGGA-GNTTCTCTTCAACATAAG 74  
DB 1 ATATGACACAATTAATATATTTGTATATCTCACACCGGAGTTCTCTTCAACATAAG 60  
QY 75 GAGTTAGAAATTAAGTAGGCATATGCTTCTATATTCAGATAAATTCATTTCGATTAA 134  
DB 61 KAGTTAGAAATTAAGTAGGCATATGCTTCTATATTCAGATAAATTCATTTCGATTAA 120  
QY 135 T 135  
DB 121 T 121

RESULT 13  
ADP86736  
ID ADF86736 standard; DNA; 121 BP.  
XX AC ADF86736;  
XX 26-FEB-2004 (first entry)  
XX Single nucleotide polymorphism detection human gene, SEQ ID No 319.  
DE human; single nucleotide polymorphism; microarray; side effect; gene; ds.  
XX Homo sapiens.  
XX JP2003235571-A.  
XX 26-AUG-2003.  
XX 12-FEB-2002; 2002JP-00034717.  
XX 12-FEB-2002; 2002JP-00034717.  
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.  
XX WPI; 2003-820454/77.  
XX Novel polynucleotide useful for detecting single nucleotide polymorphisms  
XX in human gene.  
PS Claim 1; SEQ ID NO 319; 704pp; Japanese.

XX The invention relates to a novel polynucleotide isolated and purified  
CC from a human gene having any one of 935 fully defined sequences as given  
CC in specification, or a sequence having a base substitution. The invention  
CC further relates to: an oligonucleotide containing single nucleotide  
CC polymorphisms; a PCR primer set chosen from the combination of two DNA  
CC fragments from any one of 1220 fully defined sequences as given in  
CC specification; a labelling probe containing the SNP containing oligo; and  
CC a microarray equipped with the SNP containing oligo. The isolated human  
CC gene of the invention is useful for detecting the single nucleotide  
CC polymorphisms in human gene. The isolated human gene is also useful for  
CC diagnosis of disease and determination of side effect to a medical agent.  
CC The isolated human gene is also effective in detecting single nucleotide  
CC polymorphisms in a human gene. This polynucleotide sequence represents  
CC one of the 935 isolated polynucleotides from a human gene of the  
CC invention.

SQ Sequence 121 BP; 26 A; 33 C; 32 G; 29 T; 0 U; 1 Other;  
Query Match 20.4%; Score 92.6; DB 10; Length 121;  
Best Local Similarity 97.9%; Pred. No. 3.7e-16;  
Matches 92; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 360 CCCACACAGGAATCTGAGCCACACAGCTGCTCGGCGGCTCCATGTGATCATC 419  
Db 1 CCCACACAGGAATCTGAGCCACACAGCTGCTCGGCGGCTCCATGTGATCATC 60

Qy 420 CTGGTCAATGAAGTGAATGTCTCTATTTCNCGGG 453  
Db 61 YTGTCATGAAGTGAATGTCTCTATTTCNCGGG 94

RESULT 14  
ID AAH35240/c  
XX AAH35240 standard; cDNA; 455 BP.  
AC AAH35240;  
DT 03-SEP-2001 (first entry)  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:2322.  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
XX Homo sapiens.  
OS  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US026524.  
XX  
PR 29-SEP-1999; 99US-0157137P.  
PR 03-NOV-1999; 99US-0163280P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
XX WPI; 2001-235357/24.  
DR  
DR P-PSDB; AAG75835.  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers.  
XX  
PS Claim 1; Page 3851; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the  
CC proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene therapy  
CC and vaccine production. N and P may be used in the prevention, diagnosis  
CC and treatment of diseases associated with inappropriate P expression. For  
CC example, N and P may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of P by expressing inactive proteins or to  
CC supplement the patient's own production of P. Additionally, N may be used  
CC to produce the colon cancer-associated P<sub>8</sub> by inserting the nucleic acids  
CC into a host cell and culturing the cell to express the proteins. N and P  
CC can be used in the prevention, diagnosis and treatment of colorectal  
CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent  
CC sequences used in the exemplification of the present invention. N.B.  
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at  
CC time of publication, meaning no sequences are present for SEQ ID NO:1027  
CC to 1052, 7921 and 7922

SQ Sequence 455 BP; 110 A; 119 C; 105 G; 106 T; 0 U; 15 Other;  
Query Match 14.2%; Score 64.4; DB 4; Length 455;

Best Local Similarity 90.5%; Pred. No. 6.2e-08;  
Matches 76; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

Qy 2 AAATCAGAGTTTAAATATGACACAATTAATAATATTTGTATATCTCACACCGGA-GNTTC 60  
Db 147 AAATCAGAGTTTAAATATGACACNCAATTAATAATATTTGTATATCTCACNCCNANGTTTC 88

Qy 61 TCTTCAACATAAGGAGTTAGAAA 84  
Db 87 TTTTCAACATAAGGAGTTAGAAA 64

RESULT 15  
ID ADO53174/c  
XX ADO53174 standard; DNA; 538 BP.  
AC ADO53174;  
DT 21-OCT-2004 (first entry)  
XX  
DE Novel canine microarray-related DNA sequence SeqID4476.  
XX  
KW canine microarray; drug screening; toxicity assay;  
KW environmental pollutant; cellular response; gene expression profile;  
KW toxic response; liver necrosis; fatty liver disease;  
KW protein adduct formation; hepatitis; dog; ds.  
XX  
OS Canis familiaris.  
XX  
PN WO2004063324-A2.  
XX  
PD 29-JUL-2004.  
XX  
PF 05-MAY-2003; 2003WO-US013853.  
XX  
PR 03-MAY-2002; 2002US-0377240P.  
XX  
XX (GENE-) GENE LOGIC INC.  
PA (PFIZ) PFIZER PROD INC.  
XX  
XX Diggans JC, Porter M, Wei T;  
XX WPI; 2004-561890/54.  
DR  
PT New isolated nucleic acid molecule, useful for drug screening and  
PT toxicity assays or for assessing the impact, including toxicity, of a  
PT compound, pharmaceutical agent or environmental pollutant on a cell or  
PT living organism.  
XX  
PS Claim 1; SEQ ID NO 4476; 41pp; English.

XX This invention is related to a novel isolated canine nucleic acid  
CC sequences and the construction of canine microarrays containing a  
CC significant portion of the canine genome. The isolated canine nucleic  
CC acid sequences of the invention may be useful for drug screening and  
CC toxicity assays. The invention is therefore useful for assessing the  
CC impact, including toxicity, of a compound, pharmaceutical agent or  
CC environmental pollutant on a cell or living organism. The methods are  
CC useful for detecting genes that are up- or down-regulated in canines in a  
CC disease state. The sequences are useful as diagnostic agents or markers  
CC to detect a cellular response in a sample individually or as part of a  
CC gene expression profile. It is also useful as a target for agents that  
CC modulate gene expression or activity. The database is useful for  
CC producing electronic Northern blots that allow the user to determine the cell  
CC type or tissue in which a given gene is expressed and to allow  
CC determination of the abundance or expression level of a given gene in a  
CC particular tissue or cell. The methods are useful for determining the  
CC similarity of a toxic response to one or more individual compounds. The  
CC methods are useful for predicting at least one toxic response or the  
CC likelihood that a compound or test agent will induce various specific  
CC pathologies such as those of the liver (liver necrosis, fatty liver  
CC disease, protein adduct formation or hepatitis), those of the kidney,  
CC heart, brain or testes, or other pathologies associated with at least one

CC of the toxins. The methods are also useful for predicting or elucidating  
CC the potential cellular pathways influenced, induced or modulated by the  
CC compound or test agent due to the similarity of the expression profile  
CC compared to the profile induced by a known toxin. The present sequence is  
CC that of a canine DNA sequence which was claimed for use during the  
CC production of a canine microarray of the invention.

XX	Sequence	538 BP; 101 A; 117 C; 110 G; 125 T; 0 U; 85 Other;
Query Match	12.0%;	Score 54.6; DB 13; Length 538;
Best Local Similarity	55.6%;	Pred. No. 3.9e-05;
Matches 135; Conservative	0;	Mismatches 99; Indels 9; Gaps 3;
QY	99	ATGCTTCCTATATTCAGATAAATTCATTTC-----GATTAAATTAATTCAGATAGAGAGA 154
Db	535	ATGNNNNNNATCTCAGATAAATTAGTTTCTATTCTATTATTAGATTCAGATAGGAGA 476
QY	155	AGTAATTTTCGGAAAGAAATGATAGCTATATTAAGCAGATATTCATTACAATACCATG 214
Db	475	AACAANNNNNGCAAAACTATAGATCTATAGAGCAGATGA----CATGACGNNNCCACT 420
QY	215	TAGAGACATAAGCAATATTTTGGCATCTCTGTCGCTCAGTAGGCCGTGTTCCCTCTG 274
Db	419	TAAAGACACAAACAGTATTTTGGCANNNNNCNTCCCCAGGCACAGGNNNNCTCCCAAG 360
QY	275	GTA-GGGCCTTTGGAGAGTACCATCTATCTAAGATGGAGGAATGCTGTGGGAAGGCGGG 333
Db	359	CTACTGGTTTGGGANNNNNNATCTATTTAGGGCGGAGAGATGTGCGCAGGAAGCAGG 300
QY	334	ATG 336
Db	299	ACG 297

Search completed: May 5, 2005, 00:37:22  
Job time : 319.957 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:46:20 ; Search time 102.218 Seconds  
(without alignments)  
7283.499 Million cell updates/sec

Title: US-08-731-499-8  
Perfect score: 455  
Sequence: 1 GAAATCAGAGTCTTAATATG.....ATTGCTCTATTTCNGGGGGT 455

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	453	99.6	455	2	US-08-680-395-8
2	453	99.6	455	4	US-08-892-695-8
3	428	94.1	4541	4	US-09-949-016-5619
4	428	94.1	767677	4	US-09-949-016-12147
5	428	94.1	767677	4	US-09-949-016-17361
6	426.4	93.7	2328	3	US-08-811-481-34
7	426.4	93.7	2328	4	US-08-876-527-34
8	414.6	91.1	4719	4	US-09-949-016-405
9	45.2	9.9	263693	4	US-09-949-016-12386
10	45.2	9.9	263694	4	US-09-949-016-16915
11	43	9.5	601	4	US-09-949-016-13582
12	43	9.5	601	4	US-09-949-016-136427
13	43	9.5	2237	1	US-08-487-1358-1
14	43	9.5	2237	2	US-08-915-972A-1
15	43	9.5	2237	2	US-09-177-909-1
16	43	9.5	3317	4	US-09-949-016-3830
17	43	9.5	3323	4	US-09-949-016-913
18	43	9.5	89220	4	US-09-949-016-12655
19	43	9.5	89224	4	US-09-949-016-13572
20	42.8	9.4	601	4	US-09-949-016-30531
21	42.8	9.4	601	4	US-09-949-016-30532
22	42.8	9.4	601	4	US-09-949-016-37150
23	42.8	9.4	601	4	US-09-949-016-37151
24	42.8	9.4	601	4	US-09-949-016-37164
25	42.8	9.4	601	4	US-09-949-016-37165
26	42.8	9.4	601	4	US-09-949-016-145868
27	42.8	9.4	601	4	US-09-949-016-145869

C 28	42.8	9.4	601	4	US-09-949-016-146136	Sequence 146136,
C 29	42.8	9.4	601	4	US-09-949-016-146137	Sequence 146137,
C 30	42.8	9.4	601	4	US-09-949-016-146404	Sequence 146404,
C 31	42.8	9.4	601	4	US-09-949-016-146405	Sequence 146405,
C 32	42.8	9.4	205044	4	US-09-949-016-15851	Sequence 15851, A
C 33	42.8	9.4	205044	4	US-09-949-016-15852	Sequence 15852, A
C 34	42.8	9.4	205044	4	US-09-949-016-15853	Sequence 15853, A
C 35	42.8	9.4	223471	4	US-09-949-016-12387	Sequence 12387, A
C 36	42.8	9.4	223471	4	US-09-949-016-12724	Sequence 12724, A
C 37	42.8	9.4	223471	4	US-09-949-016-12725	Sequence 12725, A
C 38	42	9.2	601	4	US-09-949-016-196365	Sequence 196365,
C 39	42	9.2	601	4	US-09-949-016-196366	Sequence 196366,
C 40	42	9.2	601	4	US-09-949-016-196367	Sequence 196367,
C 41	42	9.2	451924	4	US-09-949-016-12896	Sequence 12896, A
C 42	42	9.2	451925	4	US-09-949-016-17305	Sequence 17305, A
C 43	41.8	9.2	68667	4	US-09-949-016-17017	Sequence 17017, A
C 44	41.6	9.1	76281	4	US-09-949-016-12708	Sequence 12708, A
C 45	40.6	8.9	157822	4	US-09-949-016-16723	Sequence 16723, A

ALIGNMENTS

RESULT 1  
US-08-680-395-8  
; Sequence 8, Application US/08680395  
; Patent No. 5892010  
; GENERAL INFORMATION:  
; APPLICANT: Gray, Joe W.  
; APPLICANT: Collins, Colin  
; APPLICANT: Hwang, Soo-in  
; APPLICANT: Godfrey, Tony  
; APPLICANT: Kowbel, David  
; APPLICANT: Rommens, Johanna  
; TITLE OF INVENTION: Genes from the 20q13 Amplicon and Their  
; TITLE OF INVENTION: Uses  
; NUMBER OF SEQUENCES: 40  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680.395  
; FILING DATE: 15-JUL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bastian, Kevin L.  
; REGISTRATION NUMBER: 34,774  
; REFERENCE/DOCKET NUMBER: 023070-0689000US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 576-0200  
; TELEFAX: (415) 576-0300  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 455 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: 1, 455  
; LOCATION: 1, 455 /note= "cDNA clone 20sa7 for a homolog  
; OTHER INFORMATION: of rat gene BEM-1"  
; OTHER INFORMATION:  
US-08-680-395-8

Query Match 99.6%; Score 453; DB 2; Length 455;  
Best Local Similarity 100.0%; Pred. No. 6.3e-134;  
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAGTTTAAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60  
DB 1 GAAATCAGAGTTTAAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60

QY 61 TCTTCAAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120  
DB 61 TCTTCAAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120

QY 121 TCCATTTTCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180  
DB 121 TCCATTTTCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180

QY 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCAT 240  
DB 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCAT 240

QY 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 300  
DB 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 300

QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360  
DB 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360

QY 361 CCACACAGGAATCTCAGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420  
DB 361 CCACACAGGAATCTCAGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420

QY 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGGT 455  
DB 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGGT 455

## RESULT 2

US-08-892-695-8  
; Sequence 8, Application US/08892695A  
; Patent No. 6808878  
; GENERAL INFORMATION:  
; APPLICANT: Gray, Joe W  
; APPLICANT: Collins, Collin  
; APPLICANT: Hwang, Soo In  
; APPLICANT: Godfrey, Tony  
; APPLICANT: Kowel, David  
; APPLICANT: Rommens, Johanna  
; TITLE OF INVENTION: GENES FROM THE 20Q13 AMPLICON AND THEIR USES  
; FILE REFERENCE: 2500.124US3  
; CURRENT APPLICATION NUMBER: US/08/892,695A  
; CURRENT FILING DATE: 1997-07-15  
; EARLIER APPLICATION NUMBER: 08/785,532  
; EARLIER FILING DATE: 1997-01-17  
; EARLIER APPLICATION NUMBER: 08/731,499  
; EARLIER FILING DATE: 1996-10-16  
; EARLIER APPLICATION NUMBER: 08/680,395  
; EARLIER FILING DATE: 1996-07-15  
; NUMBER OF SEQ ID NOS: 59  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 455  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:20sa7  
; FEATURE:  
; NAME/KEY: modified\_base  
; LOCATION: (57)  
; OTHER INFORMATION: n is A, C, G, T, or U  
; FEATURE:  
; NAME/KEY: modified\_base

LOCATION: (449)  
US-08-892-695-8

Query Match 99.6%; Score 453; DB 4; Length 455;  
Best Local Similarity 100.0%; Pred. No. 6.3e-134;  
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAGTTTAAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60  
DB 1 GAAATCAGAGTTTAAATATGACACAATTAATATTTGTATATCTCACACCGGAGNTTC 60

QY 61 TCTTCAAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120  
DB 61 TCTTCAAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAAA 120

QY 121 TCCATTTTCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180  
DB 121 TCCATTTTCGATTAATTAATTCAGATAGAGAGAAGTAATTTTCGGAAAAGAAATGATAG 180

QY 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCAT 240  
DB 181 CTATATTAAGCAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCAT 240

QY 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 300  
DB 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 300

QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360  
DB 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGAGTGGAGTGGTCTTCTACGCTGAACC 360

QY 361 CCACACAGGAATCTCAGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420  
DB 361 CCACACAGGAATCTCAGCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420

QY 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGGT 455  
DB 421 TGGTCAATGAAGTGAATTCCTATTTTCNGGGGT 455

## RESULT 3

US-09-949-016-5619/c  
; Sequence 5619, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5619  
; LENGTH: 4541  
; TYPE: DNA  
; ORGANISM: Human  
; US-09-949-016-5619

Query Match 94.1%; Score 428; DB 4; Length 4541;  
Best Local Similarity 99.1%; Pred. No. 2.1e-125;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAAATATGACACAATTAATATTTGTATATCTCACACCGGA-GNTT 59  
DB 4541 GAAATCAGAGTTTAAATATGACACAATTAATATTTGTATATCTCACACCGGAGTTT 4482

QY 60 CTCCTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCCTATATTTCAGATAA 119  
Db 4481 CTCCTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCCTATATTTCAGATAA 4422  
QY 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAGAAATGATA 179  
Db 4421 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAGAAATGATA 4363  
QY 180 GCTATATTAAAGCAGATATTCAATCAATACCATGTAGAGACATAAGCAATATTTTGGCA 239  
Db 4362 GCTATATTAAAGCAGATATTCAATCAATACCATGTAGAGACATAAGCAATATTTTGGCA 4303  
QY 240 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299  
Db 4302 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 4243  
QY 300 ATCTAAGATGAGGAATCTGTGGAGGGCGGATGGAGTGCGTTTCTACGCTGAAC 359  
Db 4242 ATCTAAGATGAGGAATCTGTGGAGGGCGGATGGAGTGCGTTTCTACGCTGAAC 4183  
QY 360 CCCACACAGGAATCTGCAGCCACACAGCTGCCTCTCGCGCCCTTCCATGTGATCATC 419  
Db 4182 CCCACACAGGAATCTGCAGCCACACAGCTGCCTCTCGCGCCCTTCCATGTGATCATC 4123  
QY 420 CTGGTCAATGAAGTGAATTTGCTTATTTTCNGGGG 453  
Db 4122 CTGGTCAATGAAGTGAATTTGCTTATTTCTGGG 4089

## RESULT 4

US-09-949-016-12147/c  
; Sequence 12147, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12147  
; LENGTH: 767677  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(767677)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-12147

Query Match 94.1%; Score 428; DB 4; Length 767677;  
Best Local Similarity 99.1%; Pred. No. 3.7e-124;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
QY 1 GAAATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCGGA-GNMT 59  
Db 765677 GAAATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCGGAGTTT 765618  
QY 60 CTCCTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCCTATATTTCAGATAA 119  
Db 765617 CTCCTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCCTATATTTCAGATAA 765558  
QY 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 179  
Db 765557 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 765499

QY 180 GCTATATTAAAGCAGATATTCAATCAATACCATGTAGAGACATAAGCAATATTTTGGCA 239  
Db 765498 GCTATATTAAAGCAGATATTCAATCAATACCATGTAGAGACATAAGCAATATTTTGGCA 765439  
QY 240 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299  
Db 765438 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 765379  
QY 300 ATCTAAGATGAGGAATCTGTGGAGGGCGGATGGAGTGCGTTTCTACGCTGAAC 359  
Db 765378 ATCTAAGATGAGGAATCTGTGGAGGGCGGATGGAGTGCGTTTCTACGCTGAAC 765319  
QY 360 CCCACACAGGAATCTGCAGCCACACAGCTGCCTCTCGCGCCCTTCCATGTGATCATC 419  
Db 765318 CCCACACAGGAATCTGCAGCCACACAGCTGCCTCTCGCGCCCTTCCATGTGATCATC 765259  
QY 420 CTGGTCAATGAAGTGAATTTGCTTATTTTCNGGGG 453  
Db 765258 CTGGTCAATGAAGTGAATTTGCTTATTTCTGGG 765225

## RESULT 5

US-09-949-016-17361/c  
; Sequence 17361, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF  
; FILE REFERENCE: CL001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; CURRENT FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 17361  
; LENGTH: 767677  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(767677)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-949-016-17361

Query Match 94.1%; Score 428; DB 4; Length 767677;  
Best Local Similarity 99.1%; Pred. No. 3.7e-124;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
QY 1 GAAATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCGGA-GNMT 59  
Db 765677 GAAATCAGAAGTTTAATATGACACAAATTAATATATTTGTATATCTCACCGGAGTTT 765618  
QY 60 CTCCTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCCTATATTTCAGATAA 119  
Db 765617 CTCCTCAACATAGGAGTTAGAAATTACAGTAGGCGATATGCTTCCTATATTTCAGATAA 765558  
QY 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 179  
Db 765557 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTTCGGAAAGAAATGATA 765499  
QY 180 GCTATATTAAAGCAGATATTCAATCAATACCATGTAGAGACATAAGCAATATTTTGGCA 239  
Db 765498 GCTATATTAAAGCAGATATTCAATCAATACCATGTAGAGACATAAGCAATATTTTGGCA 765439  
QY 240 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299  
Db 765439 TCATTTCTCGCTCAGTAGGCGGTTCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 765379

```

Db 765438 TCATTCTGCGCTCAGTAGCGCGTGTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 765379
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTAGCTGAAC 359
Db 765378 AICTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTAGCTGAAC 765319
QY 360 CCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 419
Db 765318 CCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 765259
QY 420 CTGGTCAATGAAGTGAATGTCTCTTTTTCNGGGG 453
Db 765258 CTGGTCAATGAAGTGAATGTCTCTTTTTCNGGGG 765225

RESULT 6
US-08-811-481-34/c
; Sequence 34, Application US/08811481
; Patent No. 6300093
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-08-811-481-34

Query Match 93.7%; Score 426.4; DB 3; Length 2328;
Best Local Similarity 98.9%; Pred. No. 4.6e-125;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNTT 59
Db 2282 GAAATCAGAGTTTAATATGACACAAATTAATATATTTGTATATCTCACACCGGAGTTT 2223
QY 60 CTCTTCAAACTAAGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119

```

```

Db 2222 CTCTTCAAACTAAGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 2163
QY 120 ATTCAATTTTCGATTAAATTAATTTCCAGATAGAGAGTAATTTTCGAAAAAGAAATGATA 179
Db 2162 ATTCAATTTTCGATTAAATTAATTTCCAGATAGAGAGTAATTTT-GGAAAAAGAAATGATA 2104
QY 180 GCTATATTTAAAGCAGATATTTCAATTACAATPACCATGTAGACACATAAGCAATATTTTGGCA 239
Db 2103 GCTATATTTAAAGCAGATATTTCAATTACAATPACCATGTAGACACATAAGCAATATTTTGGCA 2044
QY 240 TCATTCTGCTCCCTCAGTAGGCCGTGTTCCCTCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299
Db 2043 TCATTCTGCTCCCTCAGTAGGCCGTGTTCCCTCTCTGCTAGGCGCTTTGGAGAGTACCATCT 1984
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTACGCTGAAC 359
Db 1983 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTACGCTGAAC 1924
QY 360 CCCACACAGGAATCTGCAGCCACACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 419
Db 1923 CCCACACAGGAATCTGCAGCCACACAGCTGCTCTGCGCGCCCTTCCATGTGATCATC 1864
QY 420 CTGGTCAATGAAGTGAATGTCTCTTTTTCNGGGG 453
Db 1863 CTGGTCAATGAAGTGAATGTCTCTTTTTCNGGGG 1830

RESULT 7
US-09-876-527-34/c
; Sequence 34, Application US/09876527
; Patent No. 6627735
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Jelinek, Laura J.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Hagopian, William A.
; APPLICANT: LaGasse, James M.
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Zymogenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna

```



```
SEQUENCE DESCRIPTION: SEQ ID NO: 34;
US-09-876-527-34
Query Match 93.7%; Score 426.4; DB 4; Length 2328;
Best Local Similarity 98.9%; Pred. No. 4.6e-125;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 1 GAAATCAGAGTTTAAATGACACAAATTAATATATATTTGTTATATCTCACACCGGA-GNVT 59
Db 2282 GAAATCAGAGTTTAAATGACACAAATTAATATATATTTGTTATATCTCACACCGGAGTTT 2223

Qy 60 CTCCTTCAACATAAGGAGTTAGAAATACAAAGTAGGAGGATGCTTCCCTATATATCAGATAA 119
Db 2222 CTCCTTCAACATAAGGAGTTAGAAATTAACAAGTAGGAGGATGCTTCCCTATATATCAGATAA 2163

Qy 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATA 179
Db 2162 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATA 2104

Qy 180 GCTATATTAAGAGAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 2103 GCTATATTAAGAGAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 2044

Qy 240 TCATTCTGCTCGCTCAGTAGGCGGTTTCCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299
Db 2043 TCATTCTGCTCGCTCAGTAGGCGGTTTCCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 1984

Qy 300 ATCTAAGATGAGGAATCTCTGGGAAGGGCGGATGGAGGTGGCTTTTCTACGCTGAAC 359
Db 1983 ATCTAAGATGAGGAATCTCTGGGAAGGGCGGATGGAGGTGGCTTTTCTACGCTGAAC 1924

Qy 360 CCCACACAGGAAATCTGCAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 419
Db 1923 CCCACACAGGAAATCTGCAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 1864

Qy 420 CTGCTCAATGAAGTGAATTCCTATTTTCNGGG 453
Db 1863 CTGCTCAATGAAGTGAATTCCTATTTTCNGGG 1830

RESULT 8
US-09-949-016-405/c
; Sequence 405, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 405
; LENGTH: 4719
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-405

Query Match 91.1%; Score 414.6; DB 4; Length 4719;
Best Local Similarity 98.2%; Pred. No. 3.9e-121;
Matches 439; Conservative 0; Mismatches 6; Indels 2; Gaps 2;

Qy 8 GAAGTTTAAATGACACAAATTAATATATTTGTTATATCTCACACCGGA-GNVTCTCTTCA 66
Db 4710 GAAGTTTAAATGACACAAATTAATATATTTGTTATATCTCACACCGGAGGTTCTCTTCA 4651

SEQUENCE DESCRIPTION: SEQ ID NO: 34;
US-09-876-527-34
Query Match 93.7%; Score 426.4; DB 4; Length 2328;
Best Local Similarity 98.9%; Pred. No. 4.6e-125;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 1 GAAATCAGAGTTTAAATGACACAAATTAATATATATTTGTTATATCTCACACCGGA-GNVT 59
Db 2282 GAAATCAGAGTTTAAATGACACAAATTAATATATATTTGTTATATCTCACACCGGAGTTT 2223

Qy 60 CTCCTTCAACATAAGGAGTTAGAAATACAAAGTAGGAGGATGCTTCCCTATATATCAGATAA 119
Db 2222 CTCCTTCAACATAAGGAGTTAGAAATTAACAAGTAGGAGGATGCTTCCCTATATATCAGATAA 2163

Qy 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATA 179
Db 2162 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATA 2104

Qy 180 GCTATATTAAGAGAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
Db 2103 GCTATATTAAGAGAGATATTCATTAACAATACCATGTAGAGACATAAGCAATATTTTGGCA 2044

Qy 240 TCATTCTGCTCGCTCAGTAGGCGGTTTCCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 299
Db 2043 TCATTCTGCTCGCTCAGTAGGCGGTTTCCCTCTGGTAGGGCCCTTTGGAGAGTACCATCT 1984

Qy 300 ATCTAAGATGAGGAATCTCTGGGAAGGGCGGATGGAGGTGGCTTTTCTACGCTGAAC 359
Db 1983 ATCTAAGATGAGGAATCTCTGGGAAGGGCGGATGGAGGTGGCTTTTCTACGCTGAAC 1924

Qy 360 CCCACACAGGAAATCTGCAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 419
Db 1923 CCCACACAGGAAATCTGCAGCCCAACAGCTGCTCTGCGCGGCTTCCATGTGATCATC 1864

Qy 420 CTGCTCAATGAAGTGAATTCCTATTTTCNGGG 453
Db 1863 CTGCTCAATGAAGTGAATTCCTATTTTCNGGG 1830

RESULT 9
US-09-949-016-12386
; Sequence 12386, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12386
; LENGTH: 263693
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12386

Query Match 9.9%; Score 45.2; DB 4; Length 263693;
Best Local Similarity 50.7%; Pred. No. 0.013;
Matches 107; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Qy 10 AGTTTAAATGACACAAATTAATATATATTTGTTATATCTCACACCGGAGNTTCTCTCAAC 69
Db 251771 AATTATATATGATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 251830

Qy 70 ATAAGGAGTTAGAAATTAACAAGTAGGAGGATGCTTCCCTATATTCAGATAAAATTCATTCG 129
Db 251831 ATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 251890

Qy 130 ATTAATTTAAATTTCCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATAGCTATATAA 189
Db 251891 ATATAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 251950

Qy 190 AGCAGATATTCATTAACAATACCATGTAGAGA 220
Db 251951 AATAATAATTTGATTACATTAATAATATATATA 251981
```

```
RESULT 10
US-09-949-016-16915
; Sequence 16915, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16915
; LENGTH: 263694
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16915

Query Match          9.9%; Score 45.2; DB 4; Length 263694;
Best Local Similarity 50.7%; Pred. No. 0.013; Indels 0; Gaps 0;
Matches 107; Conservative 0; Mismatches 104;

QY 10 AGTTAATATGACAAATTAATATATATTTGTTATATCTCACACGGAGNTTCTCTCAAC 69
DB 251771 AATTATATGATTAAATTTAATTTAATATATATATATATATTTAATTTAATATATAT 251830

QY 70 ATAAGAGGTAGAAATACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTGCG 129
DB 251831 ATTATATTTAAATTAATATATATATATTTAAATTTAATATATATATATATTTAATTT 251890

QY 130 ATTAATTAATCCAGATAGAGAGTAAATTTTCGAAAGAAATGATAGCTATATATAA 189
DB 251891 ATATAATTAATTAATTAATATATATATATATATATTAATTAATTAATTAATTAATTA 251950

QY 190 AGCAGATATTCATTACAAATACCATGTGAGAGA 220
DB 251951 AATATAATTTGATTACATTAAATTAATATATA 251981

RESULT 11
US-09-949-016-35582/c
; Sequence 35582, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35582
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-35582

Query Match          9.9%; Score 45.2; DB 4; Length 263694;
Best Local Similarity 50.7%; Pred. No. 0.013; Indels 0; Gaps 0;
Matches 107; Conservative 0; Mismatches 104;

QY 10 AGTTAATATGACAAATTAATATATATTTGTTATATCTCACACGGAGNTTCTCTCAAC 69
DB 251771 AATTATATGATTAAATTTAATTTAATATATATATATATATTTAATTTAATATATAT 251830

QY 70 ATAAGAGGTAGAAATACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTGCG 129
DB 251831 ATTATATTTAAATTAATATATATATATTTAAATTTAATATATATATATATTTAATTT 251890

QY 130 ATTAATTAATCCAGATAGAGAGTAAATTTTCGAAAGAAATGATAGCTATATATAA 189
DB 251891 ATATAATTAATTAATTAATATATATATATATATATTAATTAATTAATTAATTAATTA 251950

QY 190 AGCAGATATTCATTACAAATACCATGTGAGAGA 220
DB 251951 AATATAATTTGATTACATTAAATTAATATATA 251981

RESULT 12
US-09-949-016-136427/c
; Sequence 136427, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 136427
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-136427

Query Match          9.5%; Score 43; DB 4; Length 601;
Best Local Similarity 56.4%; Pred. No. 0.0021; Indels 0; Gaps 0;
Matches 79; Conservative 0; Mismatches 61;

QY 24 CAATTAAATATATTTGTTATATCTCACACGGAGNTTCTCTCAACATAGAGTTAGAA 83
DB 276 CAATTTTATGTTATTTGAATATATCAGCAAAATTTCCATAATATATCAATTAATTTGTA 217

QY 84 ATTACAAGTAGGCATATGCTTCTATATTCAGATAAAATTCATTTCGATTAAATTAATTC 143
DB 216 ACCACATCCAGTGTCAATGCTTACTCTTAGAGTTTCAGATGAATTTCTTAAATTAATAA 157

QY 144 AGATAGAGAGAGTAATTTT 163
DB 156 AACTCCATAGTACTAATTTT 137

RESULT 13
US-08-487-135B-1
; Sequence 1, Application US/08487135B
; Patent No. 5821122
; GENERAL INFORMATION:
; APPLICANT: Yannick Guillaux; Francine Jotereau;
; APPLICANT: Thierry Boon-Falleur; Sophie Lucas;
; APPLICANT: Vincent Brichard
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES PEPTIDES
; TITLE OF INVENTION: WHICH FORM COMPLEXES WITH MHC MOLECULE HLA-A2
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felte & Lynch
```

STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,135B

FILING DATE: 07-Jun-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,135

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5821122man D

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5388

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2237

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-487-135B-1

Query Match 9.5%; Score 43; DB 1; Length 2237;

Best Local Similarity 56.4%; Pred. No. 0.0045;

Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 24 CAATTAATATATTGTATATCTCACACGGAGNTTCTTCAACATAGAGTTAGAA 83

DB 1388 CAATTTATGTATTGTAATATCAGCAAAATGAAATTTCCATAATTATCAATTAATTTGTA 1447

QY 84 ATTACAAGTAGCATATGCTTCTATATTCAGATAAATTCATTCGATTAATTAATTC 143

DB 1448 ACCACATCCAGTGTCTACTCTTAGAGTTAGATGAATCTTAAATTAATAAAAAA 1507

QY 144 AGATAGAGAGAGTAATTTT 163

DB 1508 AAGTCCATAGTACTAATTTT 1527

#### RESULT 14

US-08-915-972A-1

Sequence 1, Application US/08915972A

Patent No. 5986145

GENERAL INFORMATION:

APPLICANT: Yannick Gulloux; Francine Jotereau;

APPLICANT: Thierry Boon-Falleur; Sophie Lucas;

APPLICANT: Vincent Brichard

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE

TITLE OF INVENTION: COMPLEXES WITH MHC MOLECULE HLA-A2 AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/915,972A

FILING DATE: August 21, 1997

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,135

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5886145man D

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5388 - JEL/NDH

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 688-9200

TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2237

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-915-972A-1

Query Match 9.5%; Score 43; DB 2; Length 2237;

Best Local Similarity 56.4%; Pred. No. 0.0045;

Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 24 CAATTAATATATTGTATATCTCACACGGAGNTTCTTCAACATAGAGTTAGAA 83

DB 1388 CAATTTATGTATTGTAATATCAGCAAAATGAAATTTCCATAATTATCAATTAATTTGTA 1447

QY 84 ATTACAAGTAGCATATGCTTCTATATTCAGATAAATTCATTCGATTAATTAATTC 143

DB 1448 ACCACATCCAGTGTCTACTCTTAGAGTTAGATGAATCTTAAATTAATAAAAAA 1507

QY 144 AGATAGAGAGAGTAATTTT 163

DB 1508 AAGTCCATAGTACTAATTTT 1527

#### RESULT 15

US-09-177-909-1

Sequence 1, Application US/09177909

Patent No. 5958711

GENERAL INFORMATION:

APPLICANT: Yannick Gulloux; Francine Jotereau;

APPLICANT: Thierry Boon-Falleur; Sophie Lucas;

APPLICANT: Vincent Brichard

TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULES PEPTIDES WHICH

TITLE OF INVENTION: FORM COMPLEXES WITH MHC MOLECULE HLA-A2 AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fulbright & Jaworski

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

COMPUTER: IBM PS/2

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/177,909

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/915,972

FILING DATE: August 21, 1997

APPLICATION NUMBER: 08/487,135

FILING DATE: 07 June 1995

ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 5958711man D

REGISTRATION NUMBER: 30,946

REFERENCE/DOCKET NUMBER: LUD 5388 - JEL/NDH  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2237  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-177-909-1

Query Match 9.5%; Score 43; DB 2; Length 2237;  
Best Local Similarity 56.4%; Pred. No. 0.0045;  
Matches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;  
QY 24 CAATTAAATATATTTGTATATCTCACACCGGAGNTTCTCTTCAACATTAAGGAGTTAGAA 83  
Db 1388 CAATTTTATGTATTTTGAATATCAGCAAAATTTCCATAAATTATCATTAATTTGTA 1447  
QY 84 ATTACAAGTAGGCATATGCTTCTCTATATTCAGATAAATTCATTTCGATTAAATTAATCC 143  
Db 1448 ACCACATCCAGTGTCTATGCTTACTCTCTTAGAGTTTCAGATGAATCTTAAATTAATAAAA 1507  
QY 144 AGATAGAGAGAAGTAATTTT 163  
Db 1508 AAGTCCATAGTACTAATTTT 1527

Search completed: May 5, 2005, 07:03:25  
Job time : 110.552 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 5, 2005, 03:59:12 ; Search time 625.302 Seconds  
(without alignments)  
4449.254 Million cell updates/sec

Title: US-08-731-499-8

Perfect score: 455

Sequence: 1 GAATCAGAACTTAAATG.....ATTGCTCTATTTCNGGGGT 455

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5654200 seqs, 3057283753 residues

Total number of hits satisfying chosen parameters: 11308400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.\*

```

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
22: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	99.6	455	8	US-08-731-499-8
2	439	96.5	1258	15	US-10-106-698-1845
3	426.4	93.7	2328	9	US-09-876-527-34
4	426.4	93.7	2328	16	US-10-124-089-34
5	426	93.6	4767	19	US-10-887-553A-435
6	254.8	56.0	282	14	US-10-060-036-1301
7	64.4	14.2	455	15	US-10-106-698-2332
8	44.2	9.7	6811	15	US-10-240-485-93
9	44.2	9.7	11422	15	US-10-311-455-192
10	44.2	9.7	11422	17	US-10-257-166-18
11	44.2	9.7	37515	18	US-10-433-793-27
					Sequence 8, Appli
					Sequence 1845, Ap
					Sequence 34, Appl
					Sequence 435, App
					Sequence 1301, Ap
					Sequence 2332, Ap
					Sequence 93, Appl
					Sequence 132, App
					Sequence 18, Appl
					Sequence 27, Appl

```

12 43.2 9.5 670 13 US-10-027-632-180643
13 43.2 9.5 670 17 US-10-027-632-180643
14 43 9.5 2815 9 US-09-925-298-217
15 43 9.5 2815 14 US-10-102-806-217
16 43 9.5 61020 17 US-10-221-714A-513
17 42.4 9.3 3673778 16 US-10-312-841-1
18 42 9.2 1132 18 US-10-653-047-5452
19 42 9.2 8056 18 US-10-473-126-386
20 42 9.2 13574 15 US-10-311-455-1290
21 41.2 9.1 6988 15 US-10-311-455-2414
22 41.2 9.1 3673778 16 US-10-312-841-2
23 40.6 8.9 794 18 US-10-363-345A-29543
24 40.6 8.9 794 18 US-10-363-345A-29544
25 40.6 8.9 794 19 US-10-363-483A-29543
26 40.6 8.9 794 19 US-10-363-483A-29544
27 40.4 8.8 10048 15 US-10-433-793-45
28 40 8.8 7589 15 US-10-240-453-263
29 40 8.8 8246 15 US-10-311-455-176
30 39.8 8.7 413 17 US-10-424-599-90506
31 39.8 8.7 6365 15 US-10-311-455-97
32 39.8 8.7 14316 17 US-10-221-613-407
33 39.6 8.7 8771 15 US-10-311-455-1797
34 39.4 8.7 6436 15 US-10-311-455-1665
35 39.2 8.6 5695 17 US-10-221-714A-260
36 39.2 8.6 17848 14 US-10-239-676-28
37 39.2 8.6 17848 15 US-10-240-453-38
38 39.2 8.6 17848 17 US-10-257-166-58
39 39 8.6 547 18 US-10-363-345A-1577
40 39 8.6 547 19 US-10-363-345A-1578
41 39 8.6 547 19 US-10-363-483A-1577
42 39 8.6 547 19 US-10-363-483A-1578
43 39 8.6 846 17 US-10-282-122A-11017
44 39 8.6 5511 15 US-10-311-455-1843
45 39 8.6 6106 15 US-10-311-455-1446

```

#### ALIGNMENTS

#### RESULT 1

```

US-08-731-499-8
; Sequence 8, Application US/08731499
; Publication No. US20030148270A1
; GENERAL INFORMATION:
; APPLICANT: GRAY, Joe W.
; APPLICANT: COLLINS, Colin
; APPLICANT: HWANG, Soo-In
; APPLICANT: GODFREY, Tony
; APPLICANT: KOWBEL, David
; APPLICANT: ROMMENS, Johanna
; TITLE OF INVENTION: GENES FROM THE 20q13 AMPLICON AND THEIR
; TITLE OF INVENTION: USES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/731,499
; FILING DATE: 16-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US 08/680,395
; FILING DATE: 15-JUL-1996
; ATTORNEY/AGENT INFORMATION:

```

```

Sequence 180643,
Sequence 180643,
Sequence 217, App
Sequence 217, App
Sequence 513, App
Sequence 1, Appli
Sequence 5452, Ap
Sequence 366, App
Sequence 1290, Ap
Sequence 2414, Ap
Sequence 2, Appli
Sequence 29543, A
Sequence 29544, A
Sequence 29543, A
Sequence 29544, A
Sequence 45, Appl
Sequence 263, App
Sequence 176, App
Sequence 90506, A
Sequence 97, Appl
Sequence 407, App
Sequence 1797, Ap
Sequence 1665, Ap
Sequence 260, App
Sequence 28, Appl
Sequence 38, Appl
Sequence 58, Appl
Sequence 1577, Ap
Sequence 1578, Ap
Sequence 1577, Ap
Sequence 1578, Ap
Sequence 11017, A
Sequence 1843, Ap
Sequence 1446, Ap

```

```
/ NAME: Hunter, Tom
/ REGISTRATION NUMBER: 38,498
/ REFERENCE/DOCKET NUMBER: 23070-068910
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 576-0200
/ TELEFAX: (415) 576-0300
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 455 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ FEATURE:
/ NAME/KEY: -
/ LOCATION: 1..455
/ OTHER INFORMATION: /note= "cDNA clone 20sa7 for a homolog
/ OTHER INFORMATION: of rat gene BEM-1"
US-08-731-499-8

Query Match          99.6%; Score 453; DB 8; Length 455;
Best Local Similarity 100.0%; Pred. No. 3.7e-107;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTGATATCTTCACACCGGAGNTTC 60
DB 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTGATATCTTCACACCGGAGNTTC 60

QY 61 TCTTCAACATTAAGGAGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAAA 120
DB 61 TCTTCAACATTAAGGAGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAAA 120

QY 121 TTCAATTCGATTAATTAATTCAGATAGAGAAGTAATTTTCGAAAGAAATGATAG 180
DB 121 TTCAATTCGATTAATTAATTCAGATAGAGAAGTAATTTTCGAAAGAAATGATAG 180

QY 181 CTATATTAAGCAGATATTCATTACCAATACCATGTAGACATATGCTTTTGGCAT 240
DB 181 CTATATTAAGCAGATATTCATTACCAATACCATGTAGACATATGCTTTTGGCAT 240

QY 241 CATTCTGCGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 300
DB 241 CATTCTGCGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 300

QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 360
DB 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 360

QY 361 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
DB 361 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420

QY 421 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGGT 455
DB 421 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGGT 455

RESULT 2
US-10-106-698-1845/c
/ Sequence 1845, Application US/10106698
/ Publication No. US20030109690A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruben et al.
/ TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
/ FILE REFERENCE: PA005PI
/ CURRENT APPLICATION NUMBER: US/10/106,698
/ CURRENT FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: PCT/US00/26524
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US 60/157,137
/ PRIOR FILING DATE: 1999-09-29
/ PRIOR APPLICATION NUMBER: US 60/163,280
/ PRIOR FILING DATE: 1999-11-03
```

```
/ NUMBER OF SEQ ID NOS: 8564
/ SOFTWARE: PatentIn Ver. 3.0
/ SEQ ID NO 1845
/ LENGTH: 1258
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1237)..(1237)
/ OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1845

Query Match          96.5%; Score 439; DB 15; Length 1258;
Best Local Similarity 99.3%; Pred. No. 2.5e-103;
Matches 450; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTGATATCTTCACACCGGAGNTTC 60
DB 1213 GAAATCAGAAGTTTAAATATGACACAAATTAATATATTTGATATCTTCACACCGGAGNTTC 1154

QY 61 TCTTCAACATTAAGGAGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAAA 120
DB 1153 TCTTCAACATTAAGGAGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAAA 1094

QY 121 TTCAATTCGATTAATTAATTCAGATAGAGAAGTAATTTTCGAAAGAAATGATAG 180
DB 1093 TTCAATTCGATTAATTAATTCAGATAGAGAAGTAATTTTCGAAAGAAATGATAG 1035

QY 181 CTATATTAAGCAGATATTCATTACCAATACCATGTAGACATATGCTTTTGGCAT 240
DB 1034 CTATATTAAGCAGATATTCATTACCAATACCATGTAGACATATGCTTTTGGCAT 975

QY 241 CATTCTGCGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 300
DB 974 CATTCTGCGCTCAGTAGCGCGTGTCCCTCTGCTGAGGCGCTTTGGAGAGTACCATCTA 915

QY 301 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 360
DB 914 TCTAAGATGGAGGAATGCTGTGGGAAGGCGGATGGAGTGCGTTTCTACGCTGAACC 855

QY 361 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 420
DB 854 CCACACAGGAAATCTGCAGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATCC 795

QY 421 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGG 453
DB 794 TGGTCAATGAAGTGAATTTGCTTATTTTCNGGGG 762

RESULT 3
US-09-876-527-34/c
/ Sequence 34, Application US/09876527
/ Patent No. US20020102616A1
/ GENERAL INFORMATION:
/ APPLICANT: Kindsvogel, Wayne
/ Jelinek, Laura J.
/ Sheppard, Paul O.
/ Hegopian, William A.
/ Lagasse, James M.
/ TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
/ NUMBER OF SEQUENCES: 34
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: ZymoGenetics, Inc.
/ STREET: 1201 Eastlake Avenue East
/ CITY: Seattle
/ STATE: WA
/ COUNTRY: USA
/ ZIP: 98102
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq for Windows Version 2.0
```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/876,527
; FILING DATE: 07-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
;
US-09-876-527-34
Query Match 93.7%; Score 426.4; DB 9; Length 2328;
Best Local Similarity 98.9%; Pred. No. 5.9e-100;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 1 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACGGA-GNWT 59
Db 2282 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACGAGGTTT 2223
Qy 60 CTCTTCAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119
Db 2222 CTCTTCAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 2163
Qy 120 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 179
Db 2162 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 2104
Qy 180 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 239
Db 2103 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 2044
Qy 240 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTTGGAGAGTACCATCT 299
Db 2043 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTTGGAGAGTACCATCT 1984
Qy 300 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 359
Db 1983 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 1924
Qy 360 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 419
Db 1923 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 1864
Qy 420 CTGGTCAATGAAGTGAATGCTTCTTTCNGGGG 453
Db 1863 CTGGTCAATGAAGTGAATGCTTCTTTCNGGGG 1830

RESULT 4
US-10-124-089-34/c
; Sequence 34, Application US/10124089
; Publication No. US2003016067A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; Jelinek, Laura J.
; Sheppard, Paul O.
; Hagopian, William A.
; LaGasse, James M.
;
; TITLE OF INVENTION: ISLET CELL ANTIGEN 1851

```

```

;
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,089
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; SEQUENCE DESCRIPTION: SEQ ID NO: 34:
;
US-10-124-089-34
Query Match 93.7%; Score 426.4; DB 16; Length 2328;
Best Local Similarity 98.9%; Pred. No. 5.9e-100;
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

Qy 1 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACGGA-GNWT 59
Db 2282 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTGTATATCTCACACGAGGTTT 2223
Qy 60 CTCTTCAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119
Db 2222 CTCTTCAAACATAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 2163
Qy 120 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 179
Db 2162 ATTCATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGGAAGAAATGATA 2104
Qy 180 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 239
Db 2103 GCTATATTAAGCAGATATTCATTAACAATACATGTAGACATAGCAATATTTTGGCA 2044
Qy 240 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTTGGAGAGTACCATCT 299
Db 2043 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTTGGAGAGTACCATCT 1984
Qy 300 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 359
Db 1983 ATCTAAGATGGAGGAATGCTGTGGAGAGGCGGGATGGAGGTGGCTTTTCTACGCTGAAC 1924
Qy 360 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 419
Db 1923 CCCACACAGGAAATCTGAGGCCACACAGCTGCTCTGCGCGCGCTTCCATGTGATCATC 1864
Qy 420 CTGGTCAATGAAGTGAATGCTTCTTTCNGGGG 453
Db 1863 CTGGTCAATGAAGTGAATGCTTCTTTCNGGGG 1830

```

RESULT 5  
US-10-887-553A-435/c  
; Sequence 435, Application US/10887553A  
; Publication No. US20050085436A1  
; GENERAL INFORMATION:  
; APPLICANT: Garza, Dan  
; APPLICANT: Li, Hao  
; TITLE OF INVENTION: Method to treat conditions associated  
; TITLE OF INVENTION: with insulin signalling dysregulation  
; FILE REFERENCE: 4-33262  
; CURRENT APPLICATION NUMBER: US/10/887,553A  
; CURRENT FILING DATE: 2004-07-08  
; PRIOR APPLICATION NUMBER: 60/485,883  
; PRIOR FILING DATE: 2003-08-07  
; NUMBER OF SEQ ID NOS: 1208  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 435  
; LENGTH: 4767  
; TYPE: DNA  
; ORGANISM: human  
US-10-887-553A-435

Query Match 93.6%; Score 426; DB 19; Length 4767;  
Best Local Similarity 99.1%; Pred. No. 1e-99;  
Matches 448; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY	3	AATCAGAACTTTAAATATGACACAAATTAATATATTTGATATATCTCACACCGGA-GNTTCT	61
DB	4758	AATCAGAAGTTAAATATGACACAAATTAATATATTTGATATATCTCACACCGAGGTTCT	4699
QY	62	CTTCAAAACATAGGAGCTTGAATAATCAAGTAGGCATATGCTTCCCTATATTCAGATAAAT	121
DB	4698	CTTCAAAACATAGGAGTTGAATAATCAAGTAGGCATATGCTTCCCTATATTCAGATAAAT	4639
QY	122	TCATTTCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGAAAAAGAAATGATAGC	181
DB	4638	TCATTTCGATTAAATTAATTCAGATAGAGAGAAGTAATTTTCGAAAAAGAAATGATAGC	4580
QY	182	TATATTAAGCAGATATTCATTACAATACGTAGAGACATAAGCAATATTTTGGCATC	241
DB	4579	TATATTAAGCAGATATTCATTACAATACGTAGAGACATAAGCAATATTTTGGCATC	4520
QY	242	ATTCTGTCGCTCAGTAGGCGGTTCCCTCTGGTAGGGCCCTTGGAGAGTACCATCTAT	301
DB	4519	ATTCTGTCGCTCAGTAGGCGGTTCCCTCTGGTAGGGCCCTTGGAGAGTACCATCTAT	4460
QY	302	CTAAGATGAGGAATGCTGTGGGAAGGGCGGATGGAGGTGCGTTTCTACGCTGAACCC	361
DB	4459	CTAAGATGAGGAATGCTGTGGGAAGGGCGGATGGAGGTGCGTTTCTACGCTGAACCC	4400
QY	362	CACACAGGAATCTGCAGGCCACACAGCTGCTTGGCGCGCTTCCATGTGATCATCT	421
DB	4399	CACACAGGAATCTGCAGGCCACACAGCTGCTTGGCGCGCTTCCATGTGATCATCT	4340
QY	422	GGTCAATGAAGTGAATTTGCTTATTCNGGGG	453
DB	4339	GGTCAATGAAGTGAATTTGCTTATTCNGGGG	4308

RESULT 6  
US-10-060-036-1301/c  
; Sequence 1301, Application US/10060036  
; Publication No. US20030073144A1  
; GENERAL INFORMATION:  
; APPLICANT: Benson, Darin R.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Persing, David H.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Jiang, Yugu  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER  
; FILE REFERENCE: 210121.566  
; CURRENT APPLICATION NUMBER: US/10/060,036  
; CURRENT FILING DATE: 2002-01-30  
; NUMBER OF SEQ ID NOS: 4560  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1301  
; LENGTH: 282  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-060-036-1301

Query Match 56.0%; Score 254.8; DB 14; Length 282;  
Best Local Similarity 98.2%; Pred. No. 6.7e-56;  
Matches 278; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY	11	GTTTAATATGACACAAATTAATATATTTGATATATCTCACACCGGA-GNTTCTCTTCAAAAC	69
DB	282	GTTTAATATGACACAAATTAATATATTTGATATATCTCACACCGAGGTTTCTCTTCAAAAC	223
QY	70	ATAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCCCTATATTCAGATAAAATTCATTTCG	129
DB	222	ATAAGTAGTTAGAAATTAACAAGTAGGCATATGCTTCCCTATATTCAGATAAAATTCATTTCG	163
QY	130	ATTAATTAATTCAGATAGAGAGAAGTAATTTTCGAAAAAGAAATGATAGCTATATTA	189
DB	162	ATTAATTAATTCAGATAGAGAGAAGTAATTTTCGAAAAAGAAATGATAGCTATATTA	104
QY	190	ACGATATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCATCATCTGTC	249
DB	103	ACGATATATTCATTACAATACCATGTAGAGACATAAGCAATATTTTGGCATCATCTGTC	44
QY	250	CGCTCAGTAGGCGGTTCCCTCTGCTGAGGGCCTTTGGAGAGT	292
DB	43	CGCTCAGTAGGCGGTTCCCTCTGCTGAGGGCCTTTGGAGAGT	1

RESULT 7  
US-10-106-698-2332/c  
; Sequence 2332, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
; FILE REFERENCE: PA005Pi  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 2332  
; LENGTH: 455  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (8)..(8)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (86)..(86)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (93)..(93)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature  
; LOCATION: (95)..(96)  
; OTHER INFORMATION: n equals a,t,g, or c  
; NAME/KEY: misc\_feature



LOCATION: (99)..(99)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (126)..(126)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (148)..(148)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (180)..(180)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (225)..(225)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (375)..(375)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (387)..(387)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (425)..(425)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (438)..(438)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (454)..(454)  
OTHER INFORMATION: n equals a,t,g, or c  
US-10-106-698-2332

Query Match 14.2%; Score 64.4; DB 15; Length 455;  
Best Local Similarity 90.5%; Pred. No. 1.8e-06;  
Matches 76; Conservative 0; Mismatches 7; Indels 1; Gaps 1;  
Qy 2 AAATCAGAGTTTATATGACACAATTAATATATTTGTATATCTCACACCGGA-GNTTC 60  
Db 147 AAATCAGAGTTTATATGACACAATTAATATATTTGTATATCTCACACCGGA-GNTTC 88  
Qy 61 TCTTCAACATGAAGGAGTTGAAA 84  
Db 87 TTTTCAACATGAAGGAGTTGAAA 64

RESULT 8  
US-10-240-485-93  
; Sequence 93, Application US/10240485  
; Publication No. US20030148327A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; TITLE OF INVENTION: Metastasis  
; FILE REFERENCE: 5013.1007  
; CURRENT APPLICATION NUMBER: US/10/240,485  
; CURRENT FILING DATE: 2002-10-02  
; PRIOR APPLICATION NUMBER: PCT/EP01/03970  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 202  
; SEQ ID NO 93  
; LENGTH: 6811  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-485-93

Query Match 9.7%; Score 44.2; DB 15; Length 6811;  
Best Local Similarity 51.5%; Pred. No. 1;  
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
Qy 11 GTTTAATATGACACAATTAATATATTTGTATATCTCACACCGGAGNTTCCTTCAACA 70  
Db 6368 GTTTAAAAAATAATATTTAAATTTGATTTTTTTTATTGAGATTAATTTTTTGAGATAGA 6427  
Qy 71 TAAGGAGTTAGAAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAAAATTCATTTCGA 130  
Db 6428 GAAGTAGAAGAATTTTGAATTTAGATTAAGTTTTTTGTTTAGGTAATTTATTGTTGT 6487  
Qy 131 TTAATTAATTCAGATAGAGAGAAGTAATTTTCGAAAAAGAAATGATAGCTATATTTAAA 190  
Db 6488 AAGAGTTATTTTAATGTGAGAGGTTAATTTTTTACGAAATGATTTGTTTATTGTAA 6547  
Qy 191 GCAGATATTCTATTA 204  
Db 6548 AACGAGGTTTATTA 6561

## RESULT 9

US-10-311-455-192/c  
; Sequence 192, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ  
; TITLE OF INVENTION: Cytosine methylation  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 192  
; LENGTH: 11422  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-192

Query Match 9.7%; Score 44.2; DB 15; Length 11422;  
Best Local Similarity 51.5%; Pred. No. 1.3;  
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;  
Qy 2 AAATCAGAGTTTAAATATGACACAATTAATATATTTGTATATCTCACACCGGAGNTTC 61  
Db 3770 AAATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 3711  
Qy 62 CTTCAACATGAAGGAGTTAGAAATTAAGTAGGCATATGCTTCCTATATTCAGATAAAT 121  
Db 3710 AATCATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 3651  
Qy 122 TCATTTGATTAATAATTCAGATAGAGAAGTAATTTTCGAAAAAGAAATCATAGC 181  
Db 3650 CTCCTTCGATTACTCAATTAATAATAATAATAATAATAATAATAATAATAATAATA 3591  
Qy 182 TATATTAAGCAGA 195  
Db 3590 AAACTACAAAATA 3577

```
RESULT 10
US-10-257-166-18/c
; Sequence 18, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 18
; LENGTH: 11422
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-18

Query Match          9.7%; Score 44.2; DB 17; Length 11422;
Best Local Similarity 51.5%; Pred. No. 1.3;
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 2 AAATCAGAAGTTTAATATGACACAATTAATATATTTTGTATATCTCACACCGAGNTTCT 61
DB 3770 AAAATAATAATAATAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATCT 3711

QY 62 CTTCAACATAGGAGTTAGAAATTACAAGTAGGCATATGCTTCTCTATATTCAGATAAAAT 121
DB 3710 AATCAATTAATAAATAATTAATAATAAATAAATAAATAAATAAATAAATAAATAAATTT 3651

QY 122 TCATTTTCGATTAAATTAATCCAGATAGAGAGAGTAATTTTCGGAAGAAAGAAATGATAGC 181
DB 3650 CTCCTTCGATTACTCAATTAACATAATAAATAAATAAATAAATAAATAAATAAATAAAT 3591

QY 182 TATATTAAGCAGA 195
DB 3590 AAAACTACAAATA 3577

RESULT 11
US-10-433-793-27
; Sequence 27, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 27
; LENGTH: 37515
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-27

Query Match          9.7%; Score 44.2; DB 18; Length 37515;
Best Local Similarity 51.5%; Pred. No. 2.2;
Matches 100; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 11 GTTTAATATGACACAATTAATAATATTTGTATATCTCACACCGAGNTTCTCTTCAACA 70
```

```
DB 6368 GTTTAAAAAATAAATAATATTTAATGATTTTATTGAGATTAAATTTTGAGATAGA 6427
QY 71 TAAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTCTATATTCAGATAAAATTCATTCCA 130
DB 6428 GAAGTAGAAAGAAATTTTGAATTAGGATTAAATAGTTTTTTGTTAGGTAATTTATTTGTGT 6487
QY 131 TTAATTAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATGATAGCTATATATAA 190
DB 6488 AAGAGTTAATTAATGCTGGAGAGGTTAATTTTACGAAATGATATGTTTATTATTGTAA 6547
QY 191 GCAGATATTCATTA 204
DB 6548 AACGAGGTTTATTA 6561

RESULT 12
US-10-027-632-180643
; Sequence 180643, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180643
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Human
; OTHER INFORMATION:
US-10-027-632-180643

Query Match          9.5%; Score 43.2; DB 13; Length 670;
Best Local Similarity 49.8%; Pred. No. 0.66;
Matches 108; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 9 AAGTTTAAATATGACACAATTAATAATATTTGTATATCTCACACCGAGNTTCTCTTCAAA 68
DB 283 AAGAGGAGAGTGGACATTTCAATTTATTTGGAAAAAATCAAGCTGGAATTTGAATATCCAA 342
QY 69 CATAAGGAGTTAGAATTTACAGTAGGCATATGCTTCTCTATATTCAGATAAAATTCATTTC 128
DB 343 TAATACCAATTTTAAAAACAATAAGTAAATCAAGACTTCTTCAGATAAAACAATGAATCCA 402
QY 129 GATTAAATTAATTCAGATAGAGAGAAAGTAATTTTCGAAAAAGAAATGATAGCTATATTA 188
DB 403 CTACCATTAATAAGTAATTTAAAAAATAAATTTGGGAAGAAAAAATAATCCATTA 462
QY 189 AAGCAGATATTCATTAACATACCATGTAGAGACATAA 225
DB 463 AAGAAGGCTCTGGAATGCAACAAGAAATACTGAGAAAA 499

RESULT 13
US-10-027-632-180643
; Sequence 180643, Application US/10027632
```

Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.

TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

POLYMORPHISMS IN THE HUMAN GENOME

FILE REFERENCE: 108827.129

CURRENT FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 180643

LENGTH: 670

TYPE: DNA

ORGANISM: Human

US-10-027-632-180643

Query Match

Best Local Similarity 9.5%; Score 43.2; DB 17; Length 670;

Mismatches 108; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 9 AAGTTTATATGACACAAATTAATATATTTCTATATCTCACACCGGAGNTTCTCTCAA 68

Db 283 AAGAGGAGAGTGGACATTTCAAATTTATGGAAAAATCAAGCTTCTTCAGATAAACAATGCAATTC 342

Qy 69 CATAGGAGTGTAGAAATACAAAGTAGGCATATGCTTCTATATTCAGATAAATTCATTC 128

Db 343 TAATACCAATTTAAACAATAGTAAATCAAGCTTCTTCAGATAAACAATGCAATTC 402

Qy 129 GATTAAATTAATTCAGATAGAGAGTAATTTTCGAAAGAAATGATAGCTATATTA 188

Db 403 CTACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 462

Qy 189 AAGCAGATATTCATTAATACCAATACCATGTAGACATAA 225

Db 463 AAGAGGTCTGGAATGCAACAAAGAAATACAGAAAA 499

RESULT 14

US-09-925-298-217

Sequence 217, Application US/09925298

Publication No. US20020039764A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103

CURRENT APPLICATION NUMBER: US/09/925,298

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05881

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 846

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 217

LENGTH: 2815

TYPE: DNA

ORGANISM: Homo sapiens

US-09-925-298-217

Query Match 9.5%; Score 43; DB 9; Length 2815;

Best Local Similarity 56.4%; Pred. No. 1.4; Mismatches 61; Indels 0; Gaps 0;

Mismatches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 24 CAATTAATAATATTTGTATATCTCACACCGGAGNTTCTCTTCAACATAAGGAGTTAGAA 83

Db 1957 CAATTTTATGTATTTGAATATCAGCAAAATTTTCCATAATTTATCAATTAATTTGTA 2016

Qy 84 ATTACAAGTAGGCATATGCTTCTTATATTCAGATAAAATTCATTTTCGATTAAATTAATCC 143

Db 2017 ACCACATCCAGTGTCTACTCTTCTTAGAGTTTCAGATGAATCTTTAAAAATTAATAAAA 2076

Qy 144 AGATAGAGAGAGTAATTTT 163

Db 2077 AACTCCATAGTACTAATTTT 2096

RESULT 15

US-10-102-806-217

Sequence 217, Application US/10102806

Publication No. US20030054421A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA103P1C1

CURRENT APPLICATION NUMBER: US/10/102,806

CURRENT FILING DATE: 2002-03-22

PRIOR APPLICATION NUMBER: 09/925,298

PRIOR FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05881

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 846

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 217

LENGTH: 2815

TYPE: DNA

ORGANISM: Homo sapiens

US-10-102-806-217

Query Match

Best Local Similarity 56.4%; Pred. No. 1.4; Mismatches 61; Indels 0; Gaps 0;

Mismatches 79; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy 24 CAATTAATAATATTTGTATATCTCACACCGGAGNTTCTCTTCAACATAAGGAGTTAGAA 83

Db 1957 CAATTTTATGTATTTGAATATCAGCAAAATTTTCCATAATTTATCAATTAATTTGTA 2016

Qy 84 ATTACAAGTAGGCATATGCTTCTTATATTCAGATAAAATTCATTTTCGATTAAATTAATCC 143

Db 2017 ACCACATCCAGTGTCTACTCTTCTTAGAGTTTCAGATGAATCTTTAAAAATTAATAAAA 2076

Qy 144 AGATAGAGAGAGTAATTTT 163

Db 2077 AACTCCATAGTACTAATTTT 2096

Search completed: May 5, 2005, 12:12:44

Job time : 635.302 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2005, 23:17:50 ; Search time 1925.91 Seconds  
(without alignments)  
8992.766 Million cell updates/sec

Title: US-08-731-499-8

Perfect score: 455

Sequence: 1 GAATCAGAGTTTAATATG.....ATTGCTCTATTTCNGGGGT 455

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: gb\_est1:\*

2: gb\_est2:\*

3: gb\_hic:\*

4: gb\_est3:\*

5: gb\_est4:\*

6: gb\_est5:\*

7: gb\_est6:\*

8: gb\_gss1:\*

9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	453	99.6	455	7	N78571 za93b11.es1
2	428	94.1	458	4	B1712448 ie08a03.x
3	428	94.1	471	1	AA661732 nv04b06.s
4	428	94.1	472	1	AI082095 oz38a05.s
5	428	94.1	478	2	BE044633 hq87e04.x
6	428	94.1	530	5	BU728730 UI-E-Cl1-
7	428	94.1	539	1	AI955290 wx32c06.x
8	428	94.1	557	5	BU069389 im21a05.x
9	428	94.1	563	1	AI828084 wk31f02.x
10	428	94.1	642	5	BQ019198 UI-H-DT1-
11	428	94.1	715	5	BM985176 UI-CF-EC1
12	428	94.1	2413	3	AL157451 Homo sapi
13	427	93.8	454	4	BM272161 ig39h05.x
14	426.4	93.7	499	6	CA775968 ip01h07.x
15	424.8	93.4	459	1	AI433898 tll7g11.x
16	424.8	93.4	461	4	BM353165 ig76e02.x
17	424.8	93.4	486	5	BQ778146 il141f01.x
18	424.8	93.4	519	1	AI425012 tg50c03.x
19	424.8	93.4	561	5	BU077543 im38b12.x
20	424.8	93.4	586	7	CK822118 ig76e02.y
21	424.8	93.4	782	6	CB850986 UI-CF-EN1
22	420.2	92.4	551	5	BX104775 BX104775
23	417.2	91.7	449	1	AA563807 nj08h09.s
24	417	91.6	447	1	AA460668 zx64c11.s

RESULT 1	LOCUS	N78571	455 bp	mRNA	linear	EST 29-MAR-1996
25	416	91.4	458	5	BQ048938	AGENCOURT
26	416	91.4	564	2	AW292950	UI-H-BW0-
27	415.2	91.3	457	1	AI457762	tj49c10.x
28	415	91.2	441	1	AI276830	ql67f01.x
29	413	90.8	522	4	BM662818	UI-E-Cl1-
30	410.6	90.2	559	5	BQ787950	il45e02.x
31	410.6	90.2	1116	1	AL572501	AL572501
32	407.6	89.6	694	1	AV721709	AV721709
33	406.4	89.3	458	4	BG059906	naf45f08.
34	405	89.0	417	1	AI082653	ox59a02.s
35	404	88.8	456	1	AI148116	qb43e06.x
36	402.6	88.5	452	5	BQ786989	il52e05.x
37	402.2	88.4	477	1	AA101715	zk95h12.8
38	398	87.5	427	2	AW662836	hi82c03.x
39	398	87.5	436	1	AI378227	tc65c09.x
40	398	87.5	458	1	AI148121	qb43e12.x
41	396.8	87.2	1906	3	CR604034	full-leng
42	395.8	87.0	1822	3	CR590102	full-leng
43	391.6	86.1	1155	1	AL573342	AL573342
44	389.8	85.7	1844	3	CR596761	full-leng
45	388	85.3	431	1	AI972872	wr44f07.x

## ALIGNMENTS

RESULT 1  
N78571  
LOCUS za93b11.es1 Soares fetal\_lung\_NDH19W Homo sapiens cDNA clone  
DEFINITION IMAGE:300093 3', mRNA sequence.  
ACCESSION N78571  
VERSION N78571.1 GI:1241272  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Travaakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
Seq primer: ml3 -40 forward  
High quality sequence stop: 390.

FEATURES  
Location/Qualifiers  
1..455  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:1245017"  
/db\_xref="taxon:9606"  
/clone="IMAGE:300093"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares fetal lung NDH19W"  
/note="Organ: lung; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'-TGTTACCACTGATGGGCGCGCAATTTTCTTTT-3'], double-stranded cDNA was selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Facima Bernaldo. This library was constructed from the same fetus as the fetal heart library, Soares fetal heart NBH19W."

## ORIGIN

```
Query Match      99.6%; Score 453; DB 7; Length 455;
Best Local Similarity 100.0%; Pred. No. 1.5e-110;
Matches 455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGAGNTTC 60
   |||||
Db 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGAGNTTC 60

QY 61 TCTTCAACATAGGAGTTAGAAATTCACAGTAGGAGGATGCTTCTATATTCAGATAAA 120
   |||||
Db 61 TCTTCAACATAGGAGTTAGAAATTCACAGTAGGAGGATGCTTCTATATTCAGATAAA 120

QY 121 TTCAATTCGATTAAATTAATTCACAGATAGAGAGAAATTTTCGAAAAGAAATGATAG 180
   |||||
Db 121 TTCAATTCGATTAAATTAATTCACAGATAGAGAGAAATTTTCGAAAAGAAATGATAG 180

QY 181 CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCAT 240
   |||||
Db 181 CTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCAT 240

QY 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCTA 300
   |||||
Db 241 CATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCTA 300

QY 301 TCTAAGATGGAGAAATGCTGTGGAAAGGCGGAGTGGAGGTGGTCTTCTACGCTGAACC 360
   |||||
Db 301 TCTAAGATGGAGAAATGCTGTGGAAAGGCGGAGTGGAGGTGGTCTTCTACGCTGAACC 360

QY 361 CCACACAGAAATCTCGACCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420
   |||||
Db 361 CCACACAGAAATCTCGACCCACACAGCTGCTCTGCGCGGCTTCCATGTGATCATCC 420

QY 421 TGGTCAATGAAGTGAATTCCTATTTCNGGGGT 455
   |||||
Db 421 TGGTCAATGAAGTGAATTCCTATTTCNGGGGT 455
```

## RESULT 2

BI712448  
LOCUS  
DEFINITION  
1e08a03.x1 HR85 islet Homo sapiens cDNA clone IMAGE:5086349 3', mRNA sequence.

BI712448

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Unpublished (2000)

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,

MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)

Seq primer: -40UP from Gibco

High quality sequence stop: 446.

## FEATURES

## source

```
1..458
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5086349"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/notes="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
NotI; Site_2: XhoI; cDNA made by oligo-dT priming. -1kb. 5'
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."
```

## ORIGIN

```
Query Match      94.1%; Score 428; DB 4; Length 458;
Best Local Similarity 99.1%; Pred. No. 7.6e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGA-GNNT 59
   |||||
Db 4 GAAATCAGAAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGAGTTT 63

QY 60 CTCCTTCAACATAAGAGAGTTAGAAATTCACAGTAGGAGTATGCTTCTATATTCAGATAA 119
   |||||
Db 64 CTCCTTCAACATAAGAGAGTTAGAAATTCACAGTAGGAGTATGCTTCTATATTCAGATAA 123

QY 120 ATTCAATTCGATTAAATTAATTCACAGATAGAGAGAAATTTTCGAAAAGAAATGATA 179
   |||||
Db 124 ATTCAATTCGATTAAATTAATTCACAGATAGAGAGAAATTTTCGAAAAGAAATGATA 182

QY 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCA 239
   |||||
Db 183 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGACATACCAATATTTTGGCA 242

QY 240 TCATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 299
   |||||
Db 243 TCATTCGTCCGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATCT 302

QY 300 ATCTAAGATGGAGGAATGCTGTGGAAAGGCGGAGTGGAGGTGGTCTTCTACGCTGAAC 359
   |||||
Db 303 ATCTAAGATGGAGGAATGCTGTGGAAAGGCGGAGTGGAGGTGGTCTTCTACGCTGAAC 362

QY 360 CCCACACAGGAAATCTGACGCCACACAGCTCCCTCTGCGCGGCTTCCATGTGATCATC 419
   |||||
Db 363 CCCACACAGGAAATCTGACGCCACACAGCTCCCTCTGCGCGGCTTCCATGTGATCATC 422

QY 420 CTGGTCAATGAAGTGAATTCCTATTTCNGGGG 453
   |||||
Db 423 CTGGTCAATGAAGTGAATTCCTATTTCNGGGG 456
```

## RESULT 3

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AA661732

LOCUS

DEFINITION

nv04b06.s1 NCI CGAP\_Pr22 Homo sapiens cDNA clone IMAGE:1219187 3', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 471)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation by: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 470.  
Location/Qualifiers  
1..471  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:121918"  
/sex="male"  
/tissue\_type="normal prostate"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Pr22"  
/note="Organ: prostate; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from normal prostate bulk tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 94.1%; Score 428; DB 1; Length 471;  
Best Local Similarity 99.1%; Pred. No. 7.6e-104;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Qy 1 GAAATCAGAAGTTTAAATATGACACAATTAATATTTTGTATATCTCATACCGGAGNTT- 59  
Db 1 GAAATCAGAAGTTTAAATATGACACAATTAATATTTTGTATATCTCATACCGGAGTTG 60  
Qy 60 CTCCTTCAACATAGAGGTTAGAAATTCACAGTAGGCGATATGCTTCTCTATATTCAGATAA 119  
Db 61 CTCCTTCAACATAGAGGTTAGAAATTCACAGTAGGCGATATGCTTCTCTATATTCAGATAA 120  
Qy 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTTCGGAAAAAGAAATGATA 179  
Db 121 ATTCATTTTCGATTAATTAATTCAGATAGAGAGAGTAATTTT-GGAAAAAGAAATGATA 179  
Qy 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239  
Db 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239  
Qy 240 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299  
Db 240 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299  
Qy 300 ATCTAAGATGAGGAATCTGTGGAGGCGGATGGAGGTGGTTTTCTACGCTGAAC 359  
Db 300 ATCTAAGATGAGGAATCTGTGGAGGCGGATGGAGGTGGTTTTCTACGCTGAAC 359  
Qy 360 CCCACACAGGAATCTGCAGGCCCAACACAGCTGCCTCTCGCGCGGCTTCCATGTGATCATC 419  
Db 360 CCCACACAGGAATCTGCAGGCCCAACACAGCTGCCTCTCGCGCGGCTTCCATGTGATCATC 419  
Qy 420 CTGGTCAATGAAGTGAATGTCCTATTTCNGGGG 453  
|||||

Db 420 CTGGTCAATGAAGTGAATGTCCTATTCTGGGG 453  
RESULT 4  
AI082095  
LOCUS  
DEFINITION  
Oz38a05.s1 Soares NhHMPu\_S1 Homo sapiens cDNA clone IMAGE:1677584  
3', mRNA sequence.  
AI082095  
VERSION  
AI082095.1 GI:3418887  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 472)  
AUTHORS  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert length: 562 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 464.  
Location/Qualifiers  
1..472  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:1677584"  
/tissue\_type="Pooled human melanocyte, fetal heart, and pregnant uterus"  
/lab\_host="DH10B"  
/clone\_lib="Soares NhHMPu S1"  
/note="Organ: mixed (see below); Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (melanocyte 2NbHM, pregnant uterus NbHPU, and fetal heart NbH19W) were mixed, and as circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 260232-265223, 340488-345479, and 484488-489479."

Query Match 94.1%; Score 428; DB 1; Length 472;  
Best Local Similarity 99.1%; Pred. No. 7.6e-104;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Qy 1 GAAATCAGAAGTTTAAATATGACACAATTAATATTTTGTATATCTCATACCGGAGNTT- 59  
Db 4 GAAATCAGAAGTTTAAATATGACACAATTAATATTTTGTATATCTCATACCGGAGTTGT 63  
Qy 60 CTCCTTCAACATAGAGGTTAGAAATTCACAGTAGGCGATATGCTTCTCTATATTCAGATAA 119  
Db 64 CTCCTTCAACATAGAGGTTAGAAATTCACAGTAGGCGATATGCTTCTCTATATTCAGATAA 123  
Qy 120 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTTCGGAAAAAGAAATGATA 179  
Db 124 ATTCATTTTCGATTAATTAATTCAGATAGAGAGTAATTTT-GGAAAAAGAAATGATA 182  
Qy 180 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239  
Db 183 GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 242  
Qy 240 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 299  
Db 243 TCATTTCTCGCTCAGTAGGCGGTGTTCCCTCTGTGTAGGCGCTTTGGAGAGTACCATCT 302  
|||||

```

QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTCTACGCTGAAC 359
DB 303 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTCTACGCTGAAC 362
QY 360 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 419
DB 363 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 422
QY 420 CTGGTCAATGAAGTGAATGTCTTCTATTCNCGGG 453
DB 423 CTGGTCAATGAAGTGAATGTCTTCTATTCNCGGG 456

RESULT 5
BE044633
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
Source
1. 478
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3126366"
/tissue_type="follicular carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NCI CGAP Thy3"
/note="Organ: thyroid; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.4 kb. Library constructed by Life Technologies."
ORIGIN
Query Match
Best Local Similarity 94.1%; Score 428; DB 2; Length 478;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 1 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTCTATATCTCACACCGGA-GNNTT 59
DB 14 GAAATCAGAGTTTAAATATGACACAAATTAATATATTTCTATATCTCACACCGAGGTTT 73
QY 60 CTCCTTCAACATAGAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 119
DB 74 CTCCTTCAACATAGAGGAGTTAGAAATTAACAAGTAGGCATATGCTTCTATATTCAGATAA 133
QY 120 ATTCAATTCGATTAATAATTCAGATAGAGAGAAAGTAATTTTCGGAAAAAGAAATGATA 179
DB 134 ATTCAATTCGATTAATAATTCAGATAGAGAGAAAGTAATTTT-GAAAAAGAAATGATA 192

```

```

QY 180 GCTATATTAAAGCAGATATTCAATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 239
DB 193 GCTATATTAAAGCAGATATTCAATTACAAATACCATGTAGAGACATAAGCAATATTTTGGCA 252
QY 240 TCATTTCTCTCCGCTCAGTAGGCCGTGTTCCCTCTGTGTAGGGCCCTTTGGAGAGTACCATCT 299
DB 253 TCATTTCTCTCCGCTCAGTAGGCCGTGTTCCCTCTGTGTAGGGCCCTTTGGAGAGTACCATCT 312
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTCTACGCTGAAC 359
DB 313 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGGTGCGTTTCTACGCTGAAC 372
QY 360 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 419
DB 373 CCCACACAGGAAATCTGCAGCCACACACAGCTGCCCTCTGCGCGCCCTTCATGTGATCATC 432
QY 420 CTGGTCAATGAAGTGAATGTCTTCTATTCNCGGG 453
DB 433 CTGGTCAATGAAGTGAATGTCTTCTATTCNCGGG 466

RESULT 6
BU728730
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
Source
1. 530
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-E-CL1-aep-b-05-0-UI"
/tissue_type="human retina"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-CL1 is a normalized cDNA library containing the following tissue(s): retina. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double

```



stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pP73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)<sub>18</sub> tail. The sequence tag for this library is CGCGG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI). TAG TISSUE=human retina TAG\_LIB=UI-E-CL1 TAG\_SEQ=CGCGG"

## ORIGIN

Query Match	94.1%;	Score 428;	DB 5;	Length 530;
Best Local Similarity	99.1%;	Pred. No. 7.8e-104;		
Matches 450;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;
Qy	1	GAATCAGAAGTTTAAATATGACACAAATTAATATATATTTGTATATCTCACACCGA-GNTT	59	
Db	19	GAATCAGAAGTTTAAATATGACACAAATTAATATATATTTGTATATCTCACACCGAGGTTT	78	
Qy	60	CTCTTCAAAACATTAAGGAGTTAGAAATTCACAGTAGGCATATGCTTCCTATATTCAGATAA	119	
Db	79	CTCTTCAAAACATTAAGGAGTTAGAAATTCACAGTAGGCATATGCTTCCTATATTCAGATAA	138	
Qy	120	ATTCAATTCGATTAAATTTAAATTCAGATAGAGAGAAGTAAATTTTCGGAAGAAATGATA	179	
Db	139	ATTCAATTCGATTAAATTTAAATTCAGATAGAGAGAAGTAAATTTT-GGAAGAAATGATA	197	
Qy	180	GCTATATTAAGCGAGATATTCAATTACAAATACCATGTTAGAGACATAAGCAATATTTTGGCA	239	
Db	198	GCTATATTAAGCGAGATATTCAATTACAAATACCATGTTAGAGACATAAGCAATATTTTGGCA	257	
Qy	240	TCATTCGTCGGCTCAGTAGGCCGTGTTCCTCTCGGTAGGGCCCTTTGGAGAGTACCATCT	299	
Db	258	TCATTCGTCGGCTCAGTAGGCCGTGTTCCTCTCGGTAGGGCCCTTTGGAGAGTACCATCT	317	
Qy	300	ATCTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGGTGGTTTTTCTPACGCTGAAC	359	
Db	318	ATCTAAGATGGAGAAATGCTGTGGGAAGGGCGGATGGAGGTGGTTTTTCTPACGCTGAAC	377	
Qy	360	CCACACAGGAAATCTCGAGCCCAACAGAGTCGCCTCTCGCGCGGCTTCCATGTGATCATC	419	
Db	378	CCACACAGGAAATCTCGAGCCCAACAGAGTCGCCTCTCGCGCGGCTTCCATGTGATCATC	437	
Qy	420	CTGTCATGAAGTGAATTTGCTATTTCNGGGG	453	
Db	438	CTGTCATGAAGTGAATTTGCTATTTCNGGGG	471	

RESULT 7	
ACCESSION	A1955290
VERSION	1
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 539)
TITLE	NCI-CPAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-r@mail.nih.gov">cgapbs-r@mail.nih.gov</a> Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christina Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome

Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
 Insert length: 2000 Std Error: 0.00  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 414.  
 Location/Qualifiers  
 1. .539  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2545354"  
 /tissue\_type="four pooled pituitary adenomas"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI\_CGAP\_Pit1"  
 /notes="Organ: brain; Vector: pCMV-SPORT6; Site: 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies."  
 FEATURES  
 source

## ORIGIN

Query Match	94.1%	Score 428;	DB 1;	Length 539;
Best Local Similarity	99.1%	Pred. No. 7.8e-104;		
Matches 450;	Conservative 0;	Mismatches 2;	Indels 2;	Gaps 2;
Qy	1	GAATCAGAGTTTAAATATGACACAATTAATAATATATTTGTATATCTCACACCGGA-GNWT	59	
Db	4	GAATCAGAGTTTAAATATGACACAATTAATAATATATTTGTATATCTCACACCGGAGTTT	63	
Qy	60	CTCTTCAACACATAAGGAGTTTGAAGATTCACAGTAGGCATATGCTTCTATATTCAGATAA	119	
Db	64	CTCTTCAACACATAAGGAGTTTGAAGATTCACAGTAGGCATATGCTTCTATATTCAGATAA	123	
Qy	120	ATTCAITTCGATTAAATTTAAATTCAGATAGAGAGAAGTAATTTTCGAAAAGAAATGATA	179	
Db	124	ATTCAITTCGATTAAATTTAAATTCAGATAGAGAGAAGTAATTTT-GGAAAAGAAATGATA	182	
Qy	180	GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATTAAGCAATATTTTGGCA	239	
Db	183	GCTATATTAAGCAGATATTCATTACAAATACCATGTAGAGACATTAAGCAATATTTTGGCA	242	
Qy	240	TCATTTCTGTCGCTCAGTAGGCGGTGTCCTCTGTGTAGGCGCTTTTGAGAGATACCATCT	299	
Db	243	TCATTTCTGTCGCTCAGTAGGCGGTGTCCTCTGTGTAGGCGCTTTTGAGAGATACCATCT	302	
Qy	300	ATCTAAGATGGAGAAATGCTGTGGGAAGCGCGGATGGAGGTGCGTTTTCTACGCTGAAC	359	
Db	303	ATCTAAGATGGAGAAATGCTGTGGGAAGCGCGGATGGAGGTGCGTTTTCTACGCTGAAC	362	
Qy	360	CCCACACAGGAAATCTGTAGCCCCACACAGCTGCTCTGCGCCGCTTCCATGTGATCATC	419	
Db	363	CCCACACAGGAAATCTGTAGCCCCACACAGCTGCTCTGCGCCGCTTCCATGTGATCATC	422	
Qy	420	CTGGTCAATGAAGTGAATTTGCTCTATTTCNGGGG	453	
Db	423	CTGGTCAATGAAGTGAATTTGCTCTATTTCGCGG	456	

RESULT 8	BU069389	557 bp	linear	EST 27-AUG-2002
LOCUS	BU069389			
DEFINITION	in21405.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6035601			
	3', mRNA sequence.			
ACCESSION	BU069389			
VERSION	BU069389.1	GI:22510578		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
REFERENCE	1 (bases 1 to 557)			
AUTHORS	Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearc,M., Brestelli,J., Gradwohl,G., Clifton,S.,			

Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,  
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R.,  
 Williams, T., Jackson, Y., and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohp.harvard.edu

TITLE  
JOURNAL  
COMMENT

Library was constructed by Dr. J. Ferrer In vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
 University Genome Sequencing Center for information on obtaining a  
 clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 470.

FEATURES  
source

1..557  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6035601"  
 /tissue\_type="insulinoma"  
 /lab\_host="PH10B (phage-resistant)"  
 /clone\_lib="Human insulinoma"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
 XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system  
 (Stratagene) by Dr. J. Ferrer, in vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue following the Washington  
 University protocol  
 (http://genome.wustl.edu/est/lambda\_protocol.shtml).  
 Please contact Hiroshi Inoue, MD/PhD for further  
 information on this library (Metabolism Division, Permutt  
 Laboratory, Washington University School of Medicine, Box  
 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this  
 is a Washington University Pancreas EST project library."

ORIGIN

Query Match 94.1%; Score 428; DB 5; Length 557;  
 Best Local Similarity 99.1%; Pred. No. 7.9e-104;  
 Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
 1 GAAATCAGAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGA-GNTT 59  
 17 GAAATCAGAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGAGTTT 76  
 60 CTCCTCAACATAGAGGATGAGAAATTAAGATGAGCATATGCTTCTATATTCAGATAA 119  
 77 CTCCTCAACATAGAGGATGAGAAATTAAGATGAGCATATGCTTCTATATTCAGATAA 136  
 120 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAATTTTCGAAAAAGAAATGATA 179  
 137 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAATTTTCGAAAAAGAAATGATA 195  
 180 GCTATATTAAGCAGATATTCATTAACAATCCATGTAGACATAAGCAATATTTGGCA 239  
 196 GCTATATTAAGCAGATATTCATTAACAATCCATGTAGACATAAGCAATATTTGGCA 255  
 240 TCATTTCTGTCGCTAGTAGGGCGTCTCCCTCTGCTAGGGCTTTGGAGAGTACCATCT 299  
 256 TCATTTCTGTCGCTAGTAGGGCGTCTCCCTCTGCTAGGGCTTTGGAGAGTACCATCT 315  
 300 ATCTAAGATGGAGGAATGCTGTGGGAGGGCGGATGGAGGTGCGTTTCTACGCTGAAC 359  
 316 ATCTAAGATGGAGGAATGCTGTGGGAGGGCGGATGGAGGTGCGTTTCTACGCTGAAC 375  
 360 CCCACACAGGAATCTGCAGCCCAACACAGCTGCTCTGCGCCGCTTCCATGTGATCATC 419  
 376 CCCACACAGGAATCTGCAGCCCAACACAGCTGCTCTGCGCCGCTTCCATGTGATCATC 435

QY 420 CTGCTCAATGAAGTGAATTCCTATTTTCNGGG 453  
 Db |||||  
 436 CTGCTCAATGAAGTGAATTCCTATTTCTGGG 469

RESULT 9

AI828084

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 563)

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BIGAP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 1234 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 482.

Location/Qualifiers

1..563

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:2413947"

/tissue\_type="anaplastic oligodendroglioma"

/lab\_host="DH10B"

/clone\_lib="NCI-CGAP\_Brn25"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTCGAGCGCGCATAGTTTCTTTTCTTTTCTTTTCTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pT73 vector.

Library is normalized, and was constructed by Bento

Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 94.1%; Score 428; DB 1; Length 563;  
 Best Local Similarity 99.1%; Pred. No. 7.9e-104;  
 Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
 1 GAAATCAGAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGA-GNTT 59  
 3 GAAATCAGAGTTTAAATATGACACAAATTAATATTTGTATATCTCACACCGGAGTTT 62  
 60 CTCCTCAACATAGAGGATGAGAAATTAAGATGAGCATATGCTTCTATATTCAGATAA 119  
 63 CTCCTCAACATAGAGGATGAGAAATTAAGATGAGCATATGCTTCTATATTCAGATAA 122  
 120 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAATTTTCGAAAAAGAAATGATA 179  
 123 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAATTTTCGAAAAAGAAATGATA 181



DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Genetics (www.regen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA=Yes.

## FEATURES

source

## Location/Qualifiers

```
1. .715
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EC1-acc-k-02-0-UI"
/tissue_type="Lung"
/lab_host="Adult and Fetal"
/lab_host="PH108 (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EC1"
/notes="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-EC1 is a normalized cDNA library containing the
following tissue(s): Normal lung from adult and from fetal
day 64, day 87, week 19 and week 42. The library was
constructed according to Bonaldo, Lemmon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into p773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
AAGTGGCTTAC.
TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
and 380-383
TAG_LIB=UI-CF-EC1
TAG_SEQ=AAGTGGCTTAC"
```

## ORIGIN

```
Query Match          94.1%; Score 428; DB 5; Length 715;
Best Local Similarity 99.1%; Pred. No. 8.2e-104;
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTTATATCTCACACCGGA-GNNTT 59
DB 19 GAAATCAGAGTTTAAATGACACAAATTAATATTTGTTATATCTCACACCGGAGTTT 78
QY 60 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAA 119
DB 79 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAA 138
QY 120 ATTCATTTTCGATTAATTAATCCAGATAGAGAGTAATTTTCGAAAAGAAATGATA 179
DB 139 ATTCATTTTCGATTAATTAATCCAGATAGAGAGTAATTTTCGAAAAGAAATGATA 197
QY 180 GCTATATTAAGCAGATATTCATTACAATACCATTGATAGACATAGCAATATTTTGCA 239
DB 198 GCTATATTAAGCAGATATTCATTACAATACCATTGATAGACATAGCAATATTTTGCA 257
QY 240 TCATTTCTGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATT 299
DB 258 TCATTTCTGCTCAGTAGGCGGTGTTCCCTCTGTTAGGCGCTTTGGAGAGTACCATT 317
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTACGCTGAAC 359
DB 318 ATCTAAGATGGAGGAATGCTGTGGGAAGCGGGATGGAGGTGCGTTTCTACGCTGAAC 377
QY 360 CCCACACAGGAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCATC 419
DB 378 CCCACACAGGAATCTGCAGCCACACACAGCTGCCTCTGCGCGCCCTTCCATGTGATCATC 437
QY 420 CTGGTCAATGAAGTGAATTTGCTCTATTTTCNGGGG 453
DB 438 CTGGTCAATGAAGTGAATTTGCTCTATTTCTGGGG 471
```

## RESULT 12

HSM802434/c

LOCUS HSM802434 2413 bp mRNA linear HTC 22-SEP-2004  
DEFINITION Homo sapiens mRNA; cDNA DKFZp761A0712 (from clone DKFZp761A0712).

ACCESSION AL157451

VERSION AL157451.1 GI:7018465

KEYWORDS HTC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2413)

AUTHORS Anorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,

Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and

Wiemann,S.

The German cDNA Consortium

Direct Submission

Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764

Neuherberg, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

Sequenced by EMBL (European Molecular Biology Laboratories, of the

Heidelberg/Germany) within the cDNA sequencing consortium of the

German Genome Project.

This clone (DKFZp761A0712) is available at the RZPD Deutsches

Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.

Please contact RZPD for ordering:

http://www.rzpd.de/cgi-bin/products/cl.cgi?cloneID=DKFZp761A0712

Further information about the clone and the sequencing project is

available at http://mips.gsf.de/projects/cdna/.

Location/Qualifiers

1. .2413

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="RZPD:DKFZp761A0712"

/db\_xref="taxon:9606"

/clone="DKFZp761A0712"

/tissue\_type="amygdala"

/clone\_lib="761 (synonym: hamy2). Vector pSport1; host

DH10B; sites NotI + SalI"

/dev\_stage="adult"

/note="protein tyrosine phosphatase, receptor type, N

polypeptide 2, N-terminus truncated"

1. .2413

/gene="DKFZp761A0712"

&lt;12. .749

/gene="DKFZp761A0712"

/codon\_start=1

/product="hypothetical protein"

/protein\_id="CAB75665.2"

/db\_xref="GI:50978458"

/db\_xref="GOA:Q9NSR5"

/db\_xref="UniProt/TREMBL:Q9NSR5"

/translation="VPKNSLAVLTVDHSRVLLKAENSHSHSDYINAGPIMDHPNRP

AVYATQGPLPATVADFMQWVSGCVIWMITPLAENGVRQCYHYWPDGSLYHIYE

VNLVSEHLWCEDFLVRSFLKNIOTNETRTVTFHFLSWYDRGVSPSSRSLLDFFRKV

NKCYGRGSCPIIVHCSDGAGRGTYVLLDMVLNKNKAKAKEIDIAATLHLDQRPGM

VOTKEQFEFALTAVAEVNAIKALPQ"

ORIGIN

Query Match 94.1%; Score 428; DB 3; Length 2413;  
Best Local Similarity 99.1%; Pred. No. 1e-103;  
Matches 450; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1 GAAATCAGAGTTTAAATGACACAAATTAATATATTTGTTATATCTCACACCGGA-GNNTT 59

DB 2397 GAAATCAGAGTTTAAATGACACAAATTAATATATTTGTTATATCTCACACCGGAGTTT 2338

QY 60 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAA 119

DB 2337 CTCCTCAACATAGAGGTTAGAAATTACAAGTAGGCATATGCTTCCTATATTCAGATAA 2278

Qy 120 ATTCATTCGATTAAATTAATTCAGATAGAGAGAGCTAATTTTCGGAAGAAATGATA 179  
 Db 2277 ATTCATTCGATTAAATTAATTCAGATAGAGAGAGCTAATTTTCGGAAGAAATGATA 2219  
 Qy 180 GCTATATTAAGACAGATATTCATTACAAATCAGATAGAGACATAAGCAATATTTTGCA 239  
 Db 2218 GCTATATTAAGACAGATATTCATTACAAATCAGATAGAGACATAAGCAATATTTTGCA 2159  
 Qy 240 TCATTCCTGCTCAGTAGAGCGGTGTTCCCTCTGCTAGGAGCGCTTTGGAGAGTACCATCT 299  
 Db 2158 TCATTCCTGCTCAGTAGAGCGGTGTTCCCTCTGCTAGGAGCGCTTTGGAGAGTACCATCT 2099  
 Qy 300 ATCTAAGATGAGGAATGCTCTGGAGAGCGGGATGGAGGTGGTCTTCTACCGCTGAAC 359  
 Db 2098 ATCTAAGATGAGGAATGCTCTGGAGAGCGGGATGGAGGTGGTCTTCTACCGCTGAAC 2039  
 Qy 360 CCCACACAGGAAATCTGCAGCGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATC 419  
 Db 2038 CCCACACAGGAAATCTGCAGCGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATC 1979  
 Qy 420 CTGTCATGAAGTGAATTCCTTATTTTCNGGGG 453  
 Db 1978 CTGTCATGAAGTGAATTCCTTATTTCTGGG 1945

RESULT 13  
 BM272161  
 LOCUS ig39h05.x1 HR85 islet Homo sapiens cDNA 3', mRNA linear EST 20-DEC-2001  
 ACCESSION BM272161  
 VERSION BM272161.1 GI:17965439  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 454)  
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,  
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R.,  
 Williams, T., Jackson, Y. and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)  
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138  
 Tel: 617-495-1812  
 Fax: 617-495-8557  
 Email: dmelton@biohpc.harvard.edu  
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on  
 obtaining a clone please contact: Dr. Hiroshi Inoue  
 (hinoue@im.wustl.edu)  
 Seq primer: -40UP from Gibco  
 High quality sequence stop: 447.  
 Location/Qualifiers  
 1..454  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="Purified pancreatic islet"  
 /lab\_host="DH10B"  
 /clone\_lib="HR85 islet"  
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:  
 Not1; Site 2: Xho1; cDNA made by oligo-dT priming.  
 Size-selected on agarose gel. Average insert size ~1kb. 5'  
 Xho1 site was destroyed after directional cloning.  
 Amplified once. Contact information: Hiroshi Inoue, MD,  
 Metabolism Div. (Alan Permutt Lab), Washington University

FEATURES  
source

1..454  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /tissue\_type="Purified pancreatic islet"  
 /lab\_host="DH10B"  
 /clone\_lib="HR85 islet"  
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:  
 Not1; Site 2: Xho1; cDNA made by oligo-dT priming.  
 Size-selected on agarose gel. Average insert size ~1kb. 5'  
 Xho1 site was destroyed after directional cloning.  
 Amplified once. Contact information: Hiroshi Inoue, MD,  
 Metabolism Div. (Alan Permutt Lab), Washington University

## ORIGIN

Query Match 93.8%; Score 427; DB 4; Length 454;  
 Best Local Similarity 99.1%; Pred. No. 1.4e-103;  
 Matches 449; Conservative 0; Mismatches 2; Indels 2; Gaps 2;  
 Qy 2 AAATCAGAAGTTAAATATGACACAAATTAATATATTTGTATATCTCACACCGGA-GNNTC 60  
 Db 1 AAATCAGAAGTTAAATATGACACAAATTAATATATTTGTATATCTCACACCGGAGTTTC 60  
 Qy 61 TCTTCAACATATAGGAGTTAGAAATTAACAGTAGGCATATGCTTCCATATATTCAGATAAA 120  
 Db 61 TCTTCAACATATAGGAGTTAGAAATTAACAGTAGGCATATGCTTCCATATATTCAGATAAA 120  
 Qy 121 TTCAATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGAAAAAGAAATGATAG 180  
 Db 121 TTCAATTTTCGATTAAATTAATTCAGATAGAGAGAAATTAATTTTCGAAAAAGAAATGATAG 179  
 Qy 181 CTATATTAAGACAGATATTCATTACAAATCAGATAGAGACATAAGCAATATTTTGGCAT 240  
 Db 180 CTATATTAAGACAGATATTCATTACAAATCAGATAGAGACATAAGCAATATTTTGGCAT 239  
 Qy 241 CATTCCTGCTCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 300  
 Db 240 CATTCCTGCTCGCTCAGTAGGCGGTGTTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCTA 299  
 Qy 301 TCTAAGATGAGGAATGCTGTGGAGAGCGGGATGGAGGTGGTCTTCTACGCTGAAC 360  
 Db 300 TCTAAGATGAGGAATGCTGTGGAGAGCGGGATGGAGGTGGTCTTCTACGCTGAAC 359  
 Qy 361 CCACACAGGAAATCTGCAGCGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATCC 420  
 Db 360 CCACACAGGAAATCTGCAGCGCCACACAGCTGCCTCTGCGCGGCTTCCATGTGATCATCC 419  
 Qy 421 TGGTCAATGAAGTGAATTCCTTATTTTCNGGGG 453  
 Db 420 TGGTCAATGAAGTGAATTCCTTATTTCTGGG 452

## RESULT 14

CA775968  
 LOCUS ip01h07.x1 HR85 islet Homo sapiens cDNA clone linear EST 03-DEC-2002  
 DEFINITION mRNA sequence.

CA775968  
 VERSION CA775968.1 GI:26013512

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 499)

## REFERENCE

## AUTHORS

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,  
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,  
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R.,  
 Williams, T., Jackson, Y. and Bowers, Y.  
 Endocrine Pancreas Consortium  
 Unpublished (2000)

## TITLE

## JOURNAL

## COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
 Endocrine Pancreas Consortium  
 Harvard University, Howard Hughes Medical Institute  
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
 MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@biohpc.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:  
 Washington University Genome Sequencing Center For information on

School of Medicine, Box 8127, 660 South Euclid Ave., St.  
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:  
 314-362-1916, Fax: 314-747-2692."

obtaining a clone please contact: Dr. Hiroshi Inoue  
(hinoue@im.wustl.edu)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 455.

## FEATURES

source  
1. .499  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6134653"  
/tissue\_type="Purified pancreatic islet"  
/lab\_host="DH10B"  
/clone\_lib="HR85 islet"  
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:  
NotI; Site 2: XhoI; cDNA made by oligo-dT priming.  
Size-selected on agarose gel. Average insert size ~1kb. 5'  
XhoI site was destroyed after directional cloning.  
Amplified once. Contact information: Hiroshi Inoue, MD,  
Metabolism Div. (Alan Permutt Lab), Washington University  
School of Medicine, Box 8127, 660 South Euclid Ave., St.  
Louis, MO 63110. E-mail: hinoue@imgate.wustl.edu, Tel:  
314-362-1916, Fax: 314-747-2692."

## ORIGIN

Query Match 93.7%; Score 426.4; DB 6; Length 499;  
Best Local Similarity 98.9%; Pred. No. 2.1e-103;  
Matches 449; Conservative 0; Mismatches 3; Indels 2; Gaps 2;  
QY 1 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGA-GNNT 59  
Db 6 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGAGTTT 65  
QY 60 CTCCTCAACATGAAGAGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAA 119  
Db 66 CTCCTCAACATGAAGAGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAA 125  
QY 120 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 179  
Db 126 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 184  
QY 180 GCTATATTAAGCAGATATTCATTAACAATACCATGTAGACATAAGCAATATTTTGGCA 239  
Db 185 GCTATATTAAGCAGATATTCATTAACAATACCATGTAGACATAAGCAATATTTTGGCA 244  
QY 240 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299  
Db 245 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 304  
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGTGCCTTTTCTACGCTGAAC 359  
Db 305 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGTGCCTTTTCTACGCTGAAC 364  
QY 360 CCACACAGAAATCTGCAGCCACACACAGTGCCTCTGCGCGCGCTTCATGTGATCATC 419  
Db 365 CCACACAGAAATCTGCAGCCACACACAGTGCCTCTGCGCGCGCTTCATGTGATCATC 424  
QY 420 CTGGTCAATGAAGTGAATTTGCTCTATTTCTGCGG 453  
Db 425 CTGGTCAATGAAGTGAATTTGCTCTATTTCTGCGG 458

## RESULT 15

AI433898  
LOCUS t117g11.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2130788 3',  
DEFINITION mRNA sequence.  
ACCESSION AI433898  
VERSION AI433898.1 GI:4292400  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 459)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

## AUTHORS

Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapsb@mail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

## JOURNAL

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/bbrp/image/image.html](http://www-bio.lnl.gov/bbrp/image/image.html)

## COMMENT

Insert Length: 2020 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 452.

## FEATURES

Location/Qualifiers

## source

1. .459  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2130788"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Kid11"  
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker, site 1: Not I; Site 2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtracive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

## ORIGIN

Query Match 93.4%; Score 424.8; DB 1; Length 459;  
Best Local Similarity 98.7%; Pred. No. 5.5e-103;  
Matches 448; Conservative 0; Mismatches 4; Indels 2; Gaps 2;  
QY 1 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGA-GNNT 59  
Db 3 GAAATCAGAAGTTTAAATGACACAAATTAATTTGTATATCTCACACCGGAGTTT 62  
QY 60 CTCCTCAACATGAAGAGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAA 119  
Db 63 CTCCTCAACATGAAGAGTTAGAAATTAACAAGTAGGCATATGCTTCCTATATTCAGATAA 122  
QY 120 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 179  
Db 123 ATTCATTTCCGATTAATTAATTCAGATAGAGAGTAAGTAATTTTCGAAAAGAAATGATA 181  
QY 180 GCTATATTAAGCAGATATTCATTAACAATACCATGTAGACATAAGCAATATTTTGGCA 239  
Db 182 GCTATATTAAGCAGATATTCATTAACAATACCATGTAGACATAAGCAATATTTTGGCA 241  
QY 240 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 299  
Db 242 TCATTTCTGCGCTCAGTAGGCGGTGCTCCCTCTGCTAGGCGCTTTGGAGAGTACCATCT 301  
QY 300 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGTGCCTTTTCTACGCTGAAC 359  
Db 302 ATCTAAGATGGAGGAATGCTGTGGGAAGGCGGGATGGAGTGCCTTTTCTACGCTGAAC 361  
QY 360 CCACACAGAAATCTGCAGCCACACACAGTGCCTCTGCGCGCGCTTCATGTGATCATC 419  
Db 362 CCACACAGAAATCTGCAGCCACACACAGTGCCTCTGCGCGCGCTTCATGTGATCATC 421  
QY 420 CTGGTCAATGAAGTGAATTTGCTCTATTTCTGCGG 453  
Db 422 CTGGTCAATGAAGTGAATTTGCTCTATTTCTGCGG 455

Search completed: May 5, 2005, 06:54:12  
Job time : 1935.24 secs

**This Page Blank (uspto)**